

# MASTER'S THESIS

## Cloud Computing voor Big Data Analytics: EAM als instrument voor bepaling impact op processen en organisatie voor Data Quality Management

Kurvers, R.

**Award date:**  
2021

[Link to publication](#)

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



# **Cloud Computing voor Big Data Analytics: EAM als instrument voor bepaling impact op processen en organisatie voor Data Quality Management**

## **Cloud Computing for Big Data Analytics: EAM as an instrument for determining impact on processes and organization for Data Quality Management**

Opleiding: Open Universiteit, faculteit Betawetenschappen  
Masteropleiding Business Process Management & IT

Programme: Open University of the Netherlands, faculty of Science  
Master of Science Business Process Management & IT

Cursus: IM0602 Voorbereiden Afstuderen BPMIT  
IM9806 Afstudeeropdracht Business Process Management & IT

Student: Roger Kurvers

Identiteitsnummer:

Datum: 11 juli 2021

Afstudeerbegeleider: dr. R. Bos

Meelezer: ir. L. Cuijpers

Derde beoordelaar: <>

Versie nummer: 1.0

Status: Definitief

## Abstract

Organisaties hechten in toenemende mate belang aan data in het realiseren van doelstellingen waarbij Big Data Analytics een belangrijke trend vormt. Wil Big Data Analytics toegevoegde waarde hebben, moet de kwaliteit van de data, middels inrichting van Data Quality Management, geborgd zijn. Deze inrichting kan uitgedrukt worden in de lagen van Enterprise Architecture. Gegeven de eisen die Big Data Analytics stelt aan de ICT infrastructuur, overwegen organisaties dit onder te brengen in de Cloud. Introductie van Cloud Computing raakt de inrichting van organisaties, uitgedrukt in een Enterprise Architecture, en kan dus ook de inrichting van Data Quality Management raken. Dit onderzoek richt zich op de rol die Enterprise Architecture Management kan spelen bij de migratie van Big Data Analytics naar de Cloud. Focus ligt daarbij op impactbepaling van deze migratie op Data Quality Management. Het onderzoek is uitgesplitst naar 4 deelgebieden waarbij per deelgebied een theoretische verkenning is uitgevoerd, resulterend in termen en onderlinge relaties, die, aan de hand van een case study, zijn getoetst op herkenbaarheid in de praktijk. Bevindingen uit de deelgebieden zijn vervolgens bij elkaar gebracht, resulterend in een lijst met aandachtspunten die van belang zijn wil Enterprise Architecture Management zinvol worden ingezet voor impactbepaling op Data Quality Management wanneer Big Data Analytics migreert naar de Cloud.

## Sleutelbegrippen

Big Data Analytics, Cloud Computing, Data Quality Management, Enterprise Architecture, Enterprise Architecture Management

## Samenvatting

Organisaties hechten in toenemende mate belang aan data in het realiseren van hun doelstellingen. Big Data Analytics vormt hierin een belangrijke trend. Wil Big Data Analytics echter toegevoegde waarde hebben voor organisaties moet de kwaliteit van de data, middels inrichting van Data Quality Management, geborgd zijn. De inrichting van Data Quality Management kan uitgedrukt worden in de lagen van een Enterprise Architecture. Gegeven de eisen die Big Data Analytics stelt aan de infrastructuur, overwegen organisaties dit onder te brengen in de Cloud. Introductie van Cloud Computing heeft invloed op de inrichting van organisaties waaronder de inrichting van Data Quality Management.

Dit onderzoek richt zich op de rol die Enterprise Architecture Management kan spelen bij de migratie van Big Data Analytics naar de Cloud. Focus ligt daarbij op impactbepaling van deze migratie op Data Quality Management.

Het onderzoek is uitgesplitst naar 4 deelgebieden waarbij aandacht wordt besteed aan de samenhang tussen Big Data Analytics en Data Quality Management, de wijze waarop Data Quality Management zich manifesteert in een Enterprise Architecture, de wijze waarop Cloud Computing voor Big Data Analytics ingrijpt op een Enterprise Architecture en het gebruik van een Enterprise Architecture Management als instrument voor impactbepaling bij migratie naar de Cloud. Voor elk van de 4 deelgebieden is een theoretische verkenning uitgevoerd o.b.v. beschikbare literatuur en vertaald naar termen en onderlinge relaties. Deze termen en relaties zijn aan de hand van een praktijkcasus getoetst op herkenbaarheid in de praktijk. Bevindingen uit de 4 deelgebieden zijn bij elkaar gebracht en vertaald naar een lijst met aandachtspunten die van belang zijn wil Enterprise Architecture Management zinvol worden ingezet voor impactbepaling op Data Quality Management wanneer Big Data Analytics migreert naar de Cloud.

Uit de resultaten van het onderzoek kan het volgende geconcludeerd worden.

Gezien de stijgende invloed van data op het realiseren van organisatiedoelstellingen wordt het belang van Data Quality Management en bijbehorende inrichting onderstreept. Uitdrukking van deze inrichting in een Enterprise Architecture hangt echter samen met de acceptatie van dit instrument binnen een organisatie en toepasbaarheid voor Data Quality Management.

Ook voor Big Data Analytics wordt het belang van Data Quality Management onderkend. Wijze waarop dit tot uitdrukking wordt gebracht in een Enterprise Architecture is echter ook hier afhankelijk van de eerder genoemde acceptatie.

Om te komen tot de juiste Cloud ondersteuning wordt het gebruik van een Cloud Service en Deployment Model, afgeleid uit organisatiedoelstellingen, onderschreven. Doorwerking naar een Enterprise Architecture is echter ook hier afhankelijk van de acceptatie van dit instrument.

Het gebruik van Enterprise Architecture Management als methode voor impactbepaling bij migratie naar de Cloud wordt onderschreven met dien verstande dat deze methode geadopteerd is binnen de organisatie en het veranderkundig aspect, rekening houdend met het verandervermogen van de organisatie, hierin een plek krijgt.

Op basis van het bovenstaande wordt geconcludeerd dat inzet van Enterprise Architecture Management voor impactbepaling op Data Quality Management afhankelijk is van de mate van acceptatie van deze methode binnen organisatie en het gebruik van een Enterprise Architecture als instrument voor weergave van de inrichting van Big Data Analytics en Data Quality Management. Benoemde aandachtspunten, zoals weergegeven in sectie 4.6.2., kunnen hierbij een handvat vormen.

Het onderzoek is uitgevoerd bij één casusorganisatie. Het verdient aanbeveling een soortgelijk onderzoek uit voeren bij meerdere organisaties. Daarnaast wordt geconstateerd dat op de 4 afzonderlijke deelgebieden nog verdiepende onderzoeken gerechtvaardigd zijn.



## Summary

Organizations endorse an increasing importance to data in realizing their objectives.

Big Data Analytics is in this respect an important trend. The added value of Big Data Analytics depends however on the quality of the data involved. Data Quality Management insures the quality of data. The manner in which Data Quality Management is imbedded in an organization can be expressed in layers of an Enterprise Architecture.

Given the high demands of Big Data Analytics on the ICT infrastructure, organizations consider the migration towards Cloud Computing. Introducing Cloud Computing has consequences for the manner in which an organization is set up and can influence the way Data Quality Management is imbedded.

This research addresses the role of Enterprise Architecture Management in the context of migrating Big Data Analytics towards the Cloud. The research is hereby focussed on determining impact on Data Quality Management.

The research is divided in 4 subareas, focussed on the cohesion between Big Data Analytics and Data Quality Management, the manner in which Data Quality Management is expressed in an Enterprise Architecture, the manner in which Cloud Computing for Big Data Analytics effects an Enterprise Architecture and the usage of Enterprise Architecture Management as an instrument for determining impact within the context of Cloud migration.

Each of the 4 subareas is submitted to a theoretical exploration, based on literature, resulting in terms and mutual relations, and an assessment of those terms and relations in practice by means of a case study.

Findings are brought together and translated into a list with points of importance when utilizing Enterprise Architecture Management as an instrument for determining impact on Data Quality Management in case of migrating Big Data Analytics towards the Cloud.

Looking at the research results the following can be concluded.

Because of the increasing influence of data in realizing objectives, organizations underline the importance of Data Quality Management and related implementation. Expressing this implementation in an Enterprise Architecture is associated with the acceptance of this instrument within an organization and the application of it within the domain of Data Quality Management. With respect to Big Data Analytics the importance of Data Quality Management is also recognized. The application of Enterprise Architecture in this context is also related to the acceptance of this instrument for this purpose.

For choosing the appropriate Cloud support the use of a Cloud Service and Cloud Deployment Model, derived from the organization objectives, is endorsed. Translating the consequences into an Enterprise Architecture depends however on the degree in which this instrumented is accepted as useful tool.

Using Enterprise Architecture Management as a method for determining impact when migrating towards the cloud is endorsed, provided that this method is adopted within the organization and the change management aspect is not underexposed.

Based on the things mentioned above, the usage of Enterprise Architecture Management for determining impact on Data Quality Management depends on the degree of acceptance of this method within an organization and the application of Enterprise Architecture as an instrument in expressing the implementation of Big Data Analytics and Data Quality Management. The points of importance, as presented in section 4.6.2, can be a guidance in this respect.

The research is conducted using one case organization. It is recommended to conduct a similar research within other organizations. Besides this it is noticed that each of the 4 subareas can be subject for further research.

## Inhoud

Abstract .....	2
Sleutelbegrippen .....	2
Samenvatting.....	3
Summary .....	4
1.   Introductie.....	7
1.1   Achtergrond.....	7
1.2   Gebiedsverkenning.....	7
1.3   Probleemstelling.....	8
1.4   Opdrachtformulering.....	8
1.5   Motivatie/relevantie .....	9
1.6   Aanpak in hoofdlijnen .....	9
2.   Theoretisch kader.....	11
2.1   Onderzoeksaanpak.....	11
2.2   Uitvoering.....	12
2.3   Resultaten en conclusies .....	13
2.4   Doel van het vervolgonderzoek.....	14
3.   Methodologie.....	15
3.1   Conceptueel ontwerp: keuze van onderzoeksmethode(n).....	15
3.2   Technisch ontwerp: uitwerking van de methode.....	16
3.3   Gegevensanalyse .....	17
3.4   Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten.....	17
4.   Resultaten.....	19
4.1   Uitvoering .....	19
4.2   Op welke wijze hangt DQM samen met BDA? .....	20
4.2.1   Inzicht vanuit een theoretisch oogpunt .....	20
4.2.2   Toetsing aan de praktijk .....	20
4.3   Hoe manifesteert DQM zich in de lagen van een EA? .....	22
4.3.1   Inzicht vanuit een theoretisch oogpunt .....	22
4.3.2   Toetsing aan de praktijk .....	23
4.4   Op welke wijze grijpt CC voor BDA in op een EA? .....	24
4.4.1   Inzicht vanuit een theoretisch oogpunt .....	24
4.4.2   Toetsing aan de praktijk .....	25
4.5   Hoe kan EAM ingezet worden voor impactbepaling bij migratie naar CC? .....	26

4.5.1	Inzicht vanuit een theoretisch oogpunt .....	26
4.5.2	Toetsing aan de praktijk .....	27
4.6	EAM als instrument voor impactbepaling op DQM bij CC voor BDA .....	28
4.6.1	Synthese vanuit een theoretisch oogpunt .....	28
4.6.2	Vertaling naar aandachtspunten en toetsing aan interviewresultaten .....	30
5.	Discussie, conclusies en aanbevelingen .....	33
5.1	Discussie – reflectie .....	33
5.2	Conclusies .....	35
5.3	Aanbevelingen voor de praktijk .....	36
5.4	Aanbevelingen voor verder onderzoek .....	37
	Bibliografie .....	38
	Bijlage I: Lijst met afkortingen .....	41
	Bijlage II: Resultaten Samenhang DQM met BDA .....	42
	Bijlage III: Resultaten wijze waarop DQM zich manifesteert in een EA .....	47
	Bijlage IV: Resultaten wijze waarop CC voor BDA ingrijpt op een EA .....	52
	Bijlage V: Resultaten inzet EAM voor impactbepaling bij migratie CC.....	55

# 1. Introductie

## 1.1 Achtergrond

Big Data Analytics (BDA) vormt een belangrijke trend die de inrichting van organisaties raakt. Deze inrichting kan uitgedrukt worden in een Enterprise Architecture (EA). Wil BDA van toegevoegde waarde zijn voor organisaties is het van belang dat de kwaliteit van de data op orde is. Data Quality Management (DQM) moet deze kwaliteit borgen. De inrichting van DQM kan worden uitgedrukt in de lagen van bovengenoemde EA. Gegeven de eisen die BDA stelt aan de infrastructuur, overwegen organisaties dit onder te brengen in de Cloud. Introductie van Cloud Computing (CC) heeft zijn invloed op de inrichting van organisaties waaronder de inrichting van DQM. Enterprise Architecture Management (EAM) is een instrument voor het beheersen van veranderingen in de inrichting van organisaties. Gegeven de veranderingen die de introductie van CC met zich meebrengt, is de vraag welke rol EAM hierin kan vervullen.

Dit onderzoek richt zich op de rol die EAM kan spelen in geval een organisatie BDA migreert naar de Cloud. Het onderzoek focust daarbij op de gevolgen voor DQM.

In bijlage I is een lijst opgenomen met veelvuldig gehanteerde afkortingen en hun betekenis.

## 1.2 Gebiedsverkenning

Organisaties hechten in toenemende mate belang aan data en daarop gebaseerde analyses voor hun bedrijfsprocessen en bijhorende besluitvorming (Al-Ruithe, M., Benkhelifa, E., Hameed, K., 2018).

De laatste jaren is sprake van een omvangrijke groei in de hoeveelheid data, zowel in volume als variëteit, hetgeen heeft geleid tot introductie van de term Big Data (Sharma, S., 2016). Voor Big Data (BD) zijn verschillende definities in omloop (Mikalef, P., Pappas, I.O., Krogstide, J., Giannakos, M., 2018). Binnen dit onderzoek definiëren we BD als type data welke wordt gekenmerkt door grote omvang (volume), brede variëteit (variety), onzekerheid in betrouwbaarheid (veracity) en een grote omloopsnelheid (velocity) met waarde voor de onderneming (value) (Taleb, I., Serhani, M.A., Rchida, D., 2018). BDA betreft het toepassen van analyses op deze data en de presentatie van de uitkomsten op een zodanige wijze dat dit tot waarde voor de business leidt (Gantz J., Reinsel, D., 2012).

Wil de data en bijbehorende analyses van toegevoegde waarde zijn voor een organisatie moet deze kwalitatief op orde zijn. Panian onderscheidt ten aanzien van datakwaliteit de volgende dimensies: compleetheid, nauwkeurigheid en integriteit (Panian, Z., 2016). Informatie en data van hoge kwaliteit zijn essentieel voor de wijze waarop bedrijven presteren (Otto, B., Yang, Y.L., Caballero, I., 2011). Randvoorwaarde is dat het DQM op een juiste wijze is ingericht. DQM focust op het plannen, voorzien, organiseren, beheren en gebruik van data van hoge kwaliteit (Weber, K., Otto, B., Osterle, H., 2009).

Opslag en analyse van BD stelt hoge eisen aan de infrastructuur van een organisatie. CC kan hier een oplossing in bieden. CC is één van de meest significante verschuivingen in moderne ICT en services voor Enterprise toepassingen en is een belangrijke krachtige architectuur geworden voor grote en complexe verwerkingen (Hashem, I.A.T., et al., 2014). NIST (National Institute of Standards and Technology ) definieert CC als *“a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction ”* (Mell, P., Grance, T., 2011).

Invoering van CC voor BDA heeft zijn weerslag op de inrichting van organisaties. EA is reeds breed geadopteerd als een aanpak voor het plannen en besturen van complexiteit en veranderingen binnen organisaties (Niemi, E., Pekkola, S., 2019) en kan van nut zijn bij deze invoering. In dit onderzoek wordt voor EA de volgende definitie gehanteerd: *“fundamental organization of an enterprise as a socio-technical system, with the principles governing its design and development”* (Ahleman F., Stettiner E., Messerschmidt, M., Legner, C., 2012).

EAM verwezenlijkt deze EA. Onder EAM wordt hier verstaan: *“A management practice that establishes, maintains, and uses a coherent set of guidelines, architecture principles and governance regimes that provide direction for and practical help with the design and development of an enterprise’s architecture in order to achieve a vision and a strategy”* (Ahleman F., Stettiner E., Messerschmidt, M., Legner, C., 2012).

### 1.3 Probleemstelling

Organisaties hechten in toenemende mate belang aan BD en het uitvoeren van analyse hierop middels toepassing van BDA in het realiseren van doelstellingen. Wil BDA succesvol worden ingezet is het zaak dat de datakwaliteit, middels inzet van DQM, geborgd is.

Gezien de eisen die BDA stelt aan de infrastructuur overwegen organisaties dit onder te brengen in de Cloud. Introductie van CC raakt de inrichting van de organisaties waaronder de inrichting van DQM. Vraag is welke aanpassingen in de DQM inrichting moeten worden doorgevoerd zodat na migratie naar de Cloud de datakwaliteit geborgd is en BDA nog altijd succesvol kan worden ingezet. EAM en EA zijn instrumenten die kunnen worden ingezet om veranderingen in organisaties beheerst door te voeren. Vraag is in hoeverre deze instrumenten ook een rol kunnen hebben bij migratie van BDA naar de Cloud waarbij we specifiek kijken naar het effect op de inrichting van DQM.

Als we dit vraagstuk projecteren op beschikbare literatuur krijgen we hier geen antwoord op. Toepassing van BD en BDA alsmede het gebruik maken van CC is wordt veelvuldig besproken maar wordt met name vanuit een technische invalshoek belicht (Mikalef, P., Pappas, I.O., Krogstie, J., Giannakos, M., 2018). Als we hier een EA in meenemen, wordt aangegeven dat CC en BD belangrijke trends vormen voor EA's (Gampfer F., Jürgens A., Müller M., Buchkremer, R., 2018) maar wordt dit beperkt tot de infrastructurele laag is. Uitzondering vormt een onderzoek naar een EA framework voor BD binnen overheidsorganisaties maar dit is dan weer gefocust op Open Linked Data (Lnenicka, M., Komarkova, J., 2018). BD hoeft zich echter niet te beperken tot Open Linked Data. Als we, aanvullend, kijken naar DQM, levert dit ook geen beeld op, ondanks dat geconstateerd wordt dat bij toepassing van BD binnen Cloud omgevingen er uitdagingen zijn in data integriteit en slechte data kwaliteit een steeds groter probleem vormt (Hashem, I.A.T., et al., 2014).

Gegeven het ontbreken van een antwoord vanuit de literatuur vormt het eerdere genoemde een interessant onderzoeksterrein.

### 1.4 Opdrachtformulering

Het onderzoek heeft tot doel inzicht te geven in de wijze waarop EAM kan worden ingezet om de impact voor DQM te bepalen wanneer CC wordt ingezet voor BDA waarbij we ons beperken tot de business laag van een EA (zie hiervoor sectie 1.5). Gekoppeld aan bovenstaand doel, luidt de centrale onderzoeksvraag:

*Op welke wijze kan Enterprise Architecture Management zinvol worden ingezet bij het bepalen van de impact op Data Quality Management, beperkt tot de Business laag van een Enterprise Architecture, wanneer Big Data Analytics migreert naar een Cloud Computing oplossing?*

Vanuit bovenstaande vraagstelling kunnen een aantal deelgebieden worden onderkend die in samenhang beschouwd kunnen worden. Door de hoofdvraag op te splitsen in deelvragen, overeenkomend met deze deelgebieden, en de beantwoording van deze deelvragen bij elkaar te brengen, wordt getracht de onderzoeksvraag te beantwoorden.

BDA vormt in de onderzoeksvraag onderwerp van migratie naar een CC oplossing. Wil BDA in zijn algemeenheid succesvol zijn, moet de kwaliteit van de betrokken data, middels DQM, geborgd zijn, zowel vóór als na migratie. In dat kader is het relevant om te kijken wat de samenhang is tussen DQM en BDA.

CC raakt de inrichting van een organisatie. In dat kader is het van belang inzicht te hebben in de wijze waarop DQM binnen de organisatie ingericht is. Daar een EA een wijze is waarop de inrichting van een organisatie kan worden uitgedrukt, is de vraag hoe de inrichting van DQM kan worden uitgedrukt in de onderdelen van zo een EA.

BDA is binnen een organisatie op een bepaalde wijze ingericht incl. maatregelen voor borging van de datakwaliteit. Dit geheel kunnen we uitdrukken in onderdelen van een EA. Als BDA migreert naar een CC oplossing is de vraag hoe dit de BDA inrichting raakt, uitgedrukt in die onderdelen van een EA.

EAM is een methode voor het beheerst doorvoeren van veranderingen in de inrichting van een organisatie. CC migratie is zo een verandering en de vraag is welke rol EAM hierin kan spelen voor het bepalen van de impact op de inrichting.

Om basis van het bovenstaande is de onderzoeksvraag vertaald naar onderliggende deelvragen:

- a. Op welke wijze hangt DQM samen met BDA?
- b. Hoe manifesteert DQM zich in de lagen van een EA?
- c. Op welke wijze grijpt CC voor BDA in op een EA?
- d. Hoe kan EAM ingezet worden voor impactbepaling wanneer een organisatie migreert naar CC?

## 1.5 Motivatie/relevantie

Zoals in de vorige paragrafen aangegeven, heeft BDA alleen toegevoegde waarde als de kwaliteit van de data op orde is en bijbehorend DQM goed is ingericht. DQM raakt alle lagen van een EA (Otto, B., Yang, Y.L., Caballero, I., 2011). Gegeven dat veel onderzoek tot nu veelal een technische invalshoek heeft (zie sectie 1.3) beperken we ons in dit onderzoek tot de business laag van een EA. Op moment dat CC wordt ingevoerd voor BDA is het zaak processen en organisatie (als onderdeel van de business laag) dusdanig in te richten dat deze in lijn zijn met de doelstellingen voor DQM. EAM kan hierin een cruciale rol spelen. Resultaten uit dit onderzoek geven inzicht in de wijze waarop EAM kan zinvol worden ingezet. Onder een zinvolle inzet wordt in deze verstaan dat de methode organisaties, die geconfronteerd worden met de invoering van CC voor BDA, zodanige handvatten biedt dat na invoering het DQM op orde is en de kwaliteit van de data geborgd is.

## 1.6 Aanpak in hoofdlijnen

Het onderzoek heeft de volgende opbouw.

De in sectie 1.4 benoemde deelvragen vertalen zich naar de volgende deelgebieden:

- Wijze waarop DQM samenhangt met BDA
- Wijze waarop DQM zich manifesteert in de diverse lagen van een EA
- Wijze waarop CC voor BDA ingrijpt op een EA
- Wijze waarop EAM kan worden ingezet voor impactbepaling wanneer een organisatie migreert naar CC

Elk deelgebied wordt verkend vanuit een theoretisch oogpunt. Naast een eerste, samenvattend, inzicht, weergegeven in hoofdstuk 2, resulteert dit in hoofdstuk 4 in een schema met termen en onderlinge relaties die het desbetreffende deelgebied omschrijven (zie secties 4.2.1, 4.3.1, 4.4.1 en 4.5.1).

Aan de hand van een praktijkcasus wordt vervolgens getoetst of deze termen en onderlinge relaties herkend worden in de praktijk (zie secties 4.2.2, 4.3.2, 4.4.2 en 4.5.2). Termen en onderlinge relaties vormen in dit kader basis voor interviewvragen en codering van interviewresultaten (zie bijlage II tot en met bijlage V).

Ter beantwoording van de onderzoeksvraag:

- a) Worden de, vanuit het theoretisch inzicht verkregen, resultaten per deelgebied bij elkaar gebracht om te komen tot een samenhangend schema van relevante termen en onderlinge relaties waarbij de focus ligt op de Business laag van EA (zie sectie 4.6.1)
- b) Wordt het samenhangend schema met termen en relaties vertaald naar aandachtspunten die, vanuit een theoretische invalshoek, van belang zijn wil EAM zinvol ingezet kunnen worden (zie sectie 4.6.2)
- c) Worden de individuele (theoretische) aandachtspunten gerelateerd aan de bevindingen uit de praktijkcasus om vast te stellen of deze aandachtspunten onderschreven worden vanuit de praktijk (zie sectie 4.6.2). Basis hiervoor vormt de resultaten op herkenbaarheid in de praktijk van die termen en onderlinge relaties die in het samenhangend schema bij punt a zijn onderkend. Deze termen en relaties hebben immers de basis gevormd voor de aandachtspunten die bij punt b zijn onderkend.
- d) Worden o.b.v. de bevindingen uit de praktijkcasus aanvullende aandachtspunten benoemd die niet hun grondslag hebben in het verkregen theoretisch inzicht (sectie 4.6.2). Deze zijn afgeleid uit termen en relaties die aanvullend zijn benoemd bij het toetsen van herkenbaarheid van de, vanuit de theorie, verkregen termen en relaties.

## 2. Theoretisch kader

### 2.1 Onderzoeksaanpak

Om de, in sectie 1.4 geformuleerde, vraagstelling te beantwoorden is het van belang inzicht te krijgen wat vanuit de literatuur bekend is omtrent de samenhang tussen DQM, BDA, CC en EA. Om in bovenstaande samenhang inzicht te verkrijgen is het volgende zoekscenario gedefinieerd, uitgedrukt in combinatie van termen die achtereenvolgens als zoekopdrachten zijn uitgevoerd, waarbij, m.u.v. de eerste zoekopdracht, telkens 1 zoekterm buiten beschouwing wordt gelaten:

Zoekopdracht	Data Quality Management	Big Data Analytics	Enterprise Architecture	Cloud Computing
A	X	X	X	X
B	X	X	X	
C	X	X		X
D	X		X	X
E		X	X	X

Figuur 1: Zoekopdrachten met zoektermen

Om te komen tot interessante/relevante artikelen zijn achtereenvolgens voor zoekopdracht A tot en met E de volgende stappen doorlopen:

1. Uitvoeren zoekopdracht gebruik makend van Google Scholar waarbij de volgende filtering/sortering is toegepast:
  - a. Artikelen sinds 2016
  - b. Sortering op relevantie
2. Op basis van resultaat uit stap 1 selecteren van de eerste 20 artikelen per uitgevoerde zoekopdracht.
3. Op basis van resultaat uit stap 2 ontdebellen van artikelen die binnen het resultaat van de zoekopdracht meerdere keren voorkomen.
4. Op basis van resultaat uit stap 3 ontdebellen van artikelen die over zoekopdrachten heen meerdere keren voorkomen.
5. Op basis van resultaat uit stap 4 selecteren van artikelen o.b.v. inhoud van de abstract waarbij de volgende criteria zijn gehanteerd:
  - a. Gehanteerde zoektermen komen allemaal voor in de abstract en/of
  - b. In abstract gevonden zoektermen en context waarin deze gebruikt worden, kunnen gekoppeld worden aan de onderkende deelvragen
6. Op basis van resultaat uit stap 5 selecteren van artikelen waarvan inhoud van de introductie en conclusie van het artikel gekoppeld kan worden aan de onderkende deelvragen
7. Op basis van resultaten uit stap 6 selecteren van artikelen waarvan resterende inhoud van het artikel gekoppeld kan worden aan de onderkende deelvragen
8. Op basis van de resultaten uit stap 7, naast de geselecteerde artikelen, selecteren van artikelen waaraan in de literatuurlijst gerefereerd wordt, voor zover deze een bijdragen leveren in de beantwoording van de deelvragen (zgn. snowballing backward)



### Toelichting gemaakte keuzes:

In bovenstaande werkwijze zitten een aantal keuzes. Onderstaand een opsomming van deze keuzes met bijbehorende argumentatie:

- Google Scholar is als zoekmachine gehanteerd vanwege de brede basis aan bronnen waaruit deze zoekmachine put. Kanttekening hierbij vormt wel dat niet altijd duidelijk is of het geleverde resultaat een peer reviewed artikel is.
- Gegeven de snelheid van ontwikkelingen op het vakgebied is er voor gekozen alleen artikelen sinds 2016 te selecteren
- Om ervoor zorgen dat zoveel mogelijk wordt aangesloten op de, in de zoekopdracht gehanteerde, termen is gesorteerd op relevantie
- Vanwege haalbaarheid is het aantal te onderzoeken artikelen beperkt tot de eerste 20 zoekresultaten. Uitzondering vormt zoekopdracht A waarbij, gegeven dat alle termen in zoekopdracht zijn meegenomen en het beperkt aantal artikelen, alle artikelen in beschouwing zijn genomen.

## 2.2 Uitvoering

Resultaten uit de, in 2.1 gedefinieerde stappen, leidt tot het volgende overzicht.

Zoekopdracht	Data Quality Management	Big Data Analytics	Enterprise Architecture	Cloud Computing		Aantal artikelen Stap 1	Aantal artikelen stap 2	Aantal artikelen Stap 3	Aantal artikelen Stap 4	Aantal raadpleegbaar	Aantal artikelen Stap 5	Aantal artikelen Stap 6	Aantal artikelen Stap 7	Gerefereerde artikelen	Aantal artikelen stap 8
A	X	X	X	X		27	27	26	26	19	11	6	4	2	6
B	X	X	X			54	20	20	6	4	2	2	2	1	3
C	X	X		X		368	20	20	19	16	7	4	4	0	4
D	X		X	X		69	20	20	9	4	1	1	0	0	0
E		X	X	X		756	20	20	19	11	9	6	4	6	10
<b>Totaal aantal</b>										<b>79</b>	<b>54</b>	<b>30</b>	<b>19</b>	<b>14</b>	<b>23</b>

Figuur 2: Resultaten uitgevoerde stappen

In bovenstaande tabel zijn per stap het aantal gevonden artikelen weergegeven. M.b.t. het resultaat gelden de volgende opmerkingen:

- Zoekopdracht is uitgevoerd op 15 oktober 2020
- Bij zoekopdracht A is 1 doublure in de zoekresultaten gevonden
- Niet alle artikelen uit de zoekopdrachten waren raadpleegbaar. Het aantal raadpleegbare artikelen is in de grijze kolom weergegeven. De raadpleegbare documenten vormen basis voor stappen 5 tot en met 7
- Totalen per stap zijn weergegeven vanaf stap 4. Vóór stap 4 kan immers sprake zijn van dubbelstellingen binnen de zoekopdracht of over de zoekopdrachten heen
- Aantal gerefereerde artikelen betreft het aantal relevante artikelen waaraan in de, uit stap 7 geresulteerde, artikelen gerefereerd wordt
- Aantal artikelen stap 8 betreft het aantal artikelen stap 7 plus het aantal gerefereerde artikelen

## 2.3 Resultaten en conclusies

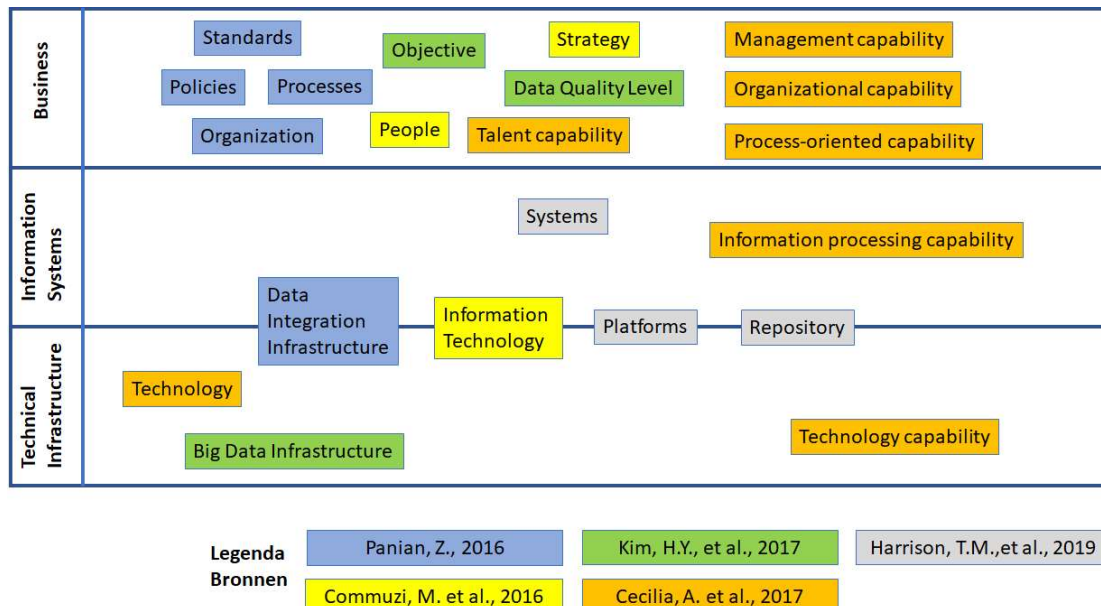
Op basis van de onderzochte artikelen ontstaat het volgende, samenvattend, inzicht. Resultaten worden verder uitgediept in hoofdstuk 4.

### *Samenhang Data Quality Management en Big Data Analytics*

Over de samenhang tussen DQM en BDA levert de literatuur het volgende, samenvattend, inzicht. Organisaties vertalen het belang van data naar data-driven business models met bijbehorende data-analyses waarin een hoge data kwaliteit een randvoorwaarde vormt (Kühne, B., Bornholt, J., Zolnowski, A., Böhm, T., 2019), (Ohbyung, K., Namyoon, L., Bongsik, S., 2014), (Popovic, A., Hackny, R., Tassabehji, R., Castelli, M., 2018). Deze kwaliteit kan uitgedrukt worden in een aantal karakteristieken die basis vormen voor maatregelen ter toetsing of aan het juiste kwaliteitsniveau wordt voldaan met bijbehorende correctieve acties (Merino, J., Caballero, I., Rivas, B., Serrano, M., Piattini, M., 2015), (Panian, Z., 2016). Dit geldt zowel voor gestructureerde als ongestructureerde data (Taleb, I., Serhani, M.A., Rchida, D., 2018) waarbij kennis inzake betekenis, structuur, integriteit, validiteit en eigenaarschap van data een vitale factor vormt om de kwaliteit te garanderen (Otto, B., 2017).

### *Data Quality Management uitgedrukt in de lagen van een Enterprise Architecture*

Over de inrichting van DQM uitgedrukt in de lagen van een EA is vanuit de literatuur maar beperkte informatie beschikbaar. In een aantal gevallen wordt EA niet expliciet benoemd maar zijn elementen uit een EA wel herkenbaar (Kim, H.Y., Cho, J-S., 2017), (Cecilia, A., Rusli, A., Rodziah, A., Yah, J., 2017), (Comuzzi, M., Patel, A., 2016), (Panian, Z., 2016) alhoewel dit ook beperkt kan zijn tot de data architectuur (Harrison, T.M, et al., 2019). In figuur 3 is getracht deze elementen alsnog te projecteren op de diverse lagen van een EA.



Figuur 3: Projectie van elementen van Data Quality Management op lagen van een Enterprise Architecture

Daarnaast zijn er artikelen waar EA wel expliciet terug komt in relatie tot BDA maar de term DQM niet wordt benoemd maar wel uit kan worden afgeleid (Gong, Y., Janssen, M., 2020), (Burmeister, F., Drews, P., Schirmer, I., 2018), (Lnenicka, M., Komarkova, J., 2018) en (Lnenicka, M., Machova, R., Komarkova, J., Cermakova, I., 2017).

Uit bovenstaande valt op te maken dat de wijze waarop DQM zich manifesteert in lagen van een EA in de literatuur niet expliciet tot uitdrukking is gebracht maar hier wel een vertaling in is te maken.

### *Impact van Cloud Computing voor Big Data Analytics op de Enterprise Architecture.*

Kijkend naar de impact van CC voor BDA op de EA leveren de artikelen inzicht in de inrichting van BDA en mogelijkheden voor CC, echter zonder aandacht voor DQM (Lnenicka, M., Komarkova, J., 2018), (Lnenicka, M., Machova, R., Komarkova, J., Cermakova, I., 2017), (Bernal, W.N., Cabalero, G.C., Sanchez, J.O., Paez-Logreira, H., 2016), (Schmidt, R., Möhring, M., 2013). Andere artikelen benoemen juist wel DQM maar beperken zich dan weer tot de processen die een rol spelen binnen BDA (Di Martino, B., Aversa, R., Cretella, G., Esposito, A., 2014), (Dong, X., Heng, H., Chao, L., Yongchuan, L., Houbo, X., 2018). Daar in deze artikelen aandacht wordt besteed aan de wijze waarop in de processen rekening wordt gehouden met maatregelen voor DQM kan dit inzicht betrokken worden in het vervolg van het onderzoek. Daarnaast wordt in een aantal gevallen de problematiek alleen vanuit een technische invalshoek belicht (Hashem, I.A.T., et al., 2014), (Demchenko, Y., Laat de, C., 2014), (Sharma, S., 2016).

### *Enterprise Architecture Management als instrument voor impactbepaling bij invoering van Cloud Computing*

Over het gebruik van EAM voor impactbepaling zijn vanuit de literatuur de inzichten te verkrijgen gericht op nut/noodzaak en implementatiewijze (Ahleman F., Stettiner E., Messerschmidt, M., Legner, C., 2012). Toegespitst op CC zien we EA als instrument benoemd (Mahmood, Z., Hill, R., 2011) maar ontbreekt het gebruik van EAM. Door bijvoorbeeld aandacht te besteden aan het ontwikkelen van een EA van de AS-IS situatie en alignment van de EA componenten met de gewenste Cloud ondersteuning in het komen tot de TO-BE situatie herkennen we wel onderdelen van een EAM.

### *Algemene conclusie*

Kijkend naar de onderzoeksvraag wordt geconcludeerd dat de resultaten uit de literatuurstudie hier slechts een gefragmenteerd beeld opleveren.

Een aantal artikelen gaan over DQM in relatie tot BDA. Alhoewel dit niet vanuit een EA perspectief wordt benaderd, zijn elementen van een EA wel herkenbaar. Daar waar EA expliciet aan de orde is in relatie tot BDA, ontbreekt het aspect van DQM volledig of wordt het slechts zijdelings benoemd. Als we vervolgens CC hierin betrekken is het inzicht wat vanuit de literatuur kan worden verkregen, zeer beperkt.

Ten aanzien van EAM in zijn algemeenheid is vanuit de literatuur het nodige beschreven. Als we dit echter koppelen aan bovenstaande thema's waarbij EAM wordt ingezet als instrument voor impactbepaling met focus op de Business laag levert de literatuur weinig tot geen inzichten. Geconstateerd wordt dat de artikelen een vertrekpunt vormen voor beantwoording van de onderzoeksvraag vanuit een theoretisch oogpunt maar hierbij de link naar de praktijk ontbreekt.

## 2.4 Doel van het vervolgonderzoek

Zoals in sectie 1.4 aangegeven, kan door een verkenning van de deelgebieden en vervolgens de resultaten hiervan bij elkaar te brengen, de onderzoeksvraag beantwoord worden. Om dit te bewerkstellingen worden de volgende stappen voor het vervolg onderkend:

- Verken de deelgebieden middels een verder verdieping vanuit de literatuur en vertaal dit per deelgebied naar termen en hun onderlinge relaties. De resultaten, weergegeven in sectie 2.3, zijn hiervoor ontoereikend en worden daarvoor in hoofdstuk 4 verder uitgewerkt.
- Toets per deelgebied of de, bij stap a onderkende, termen en hun onderlinge relaties herkend worden in de praktijk
- Beschouw de termen en onderlinge relaties, zoals onderkend bij stap a, in de context van de onderzoeksvraag en breng deze vanuit een theoretisch oogpunt samen, resulterend in een lijst met aandachtspunten, wil EAM zinvol ingezet kunnen worden.
- Vergelijk de aandachtspunten uit stap c met de bevindingen uit de praktijkcasus, uit stap b, om vast te stellen of de aandachtspunten onderschreven worden vanuit de praktijk

### 3. Methodologie

#### 3.1 Conceptueel ontwerp: keuze van onderzoeksmethode(n)

Deze sectie behandelt de methodologische keuze, doelstelling gerelateerd aan het ontwerp, de onderzoeksstrategie en de keuze omtrent de tijdshorizon die gelden voor het onderzoek.

Het doel van het onderzoek betreft het vaststellen van de wijze waarop EAM zinvol kan worden ingezet voor impactbepaling op DQM wanneer BDA migreert naar een CC oplossing.

Hierbij is gekozen voor een benadering waarbij de onderzoeksvraag is uitgesplitst naar een aantal deelgebieden waarvan de resultaten vervolgens bij elkaar worden gebracht ter beantwoording van de onderzoeksvraag.

Elk deelgebied wordt hierbij verkend vanuit een theoretisch oogpunt, resulterend in termen en onderlinge relaties, die vervolgens getoetst worden op herkenbaarheid in de praktijk.

De onderzoeksvraag wordt vervolgens ook beantwoord vanuit zowel een theoretisch als praktisch oogpunt.

Vanuit een theoretisch oogpunt worden de relevante termen en onderlinge relaties uit de deelgebieden in de context van de onderzoeksvraag bij elkaar gebracht en vertaald naar aandachtspunten die vanuit de theorie gelden, wil EAM zinvol worden ingezet.

Vanuit een praktisch oogpunt worden de bevindingen uit de praktijkcasus (ontstaan bij de toets op herkenbaarheid van termen en relaties binnen de deelgebieden) gerelateerd aan de theoretische aandachtspunten, teneinde vast te stellen of deze aandachtspunten ook onderkend worden vanuit de praktijk. De aandachtspunten hebben immers hun basis in de, binnen de deelgebieden onderkende, termen en onderlinge relaties. Mate waarin een aandachtspunt onderkend wordt vanuit de praktijk is dus afhankelijk van de mate van herkenbaarheid van onderliggende termen en relaties.

##### *Methodologische keuze*

Voor een onderzoek kan gekozen worden tussen een kwantitatieve methode, een kwalitatieve methode of een gemengde vorm (Saunders, M.K.N, Lewis, P., Thornhill, A., 2019).

Gegeven het exploratieve karakter van het onderzoek, zijnde het, vanuit deelgebieden, ontwikkelen van een lijst met aandachtspunten o.b.v. literatuur en toetsing hiervan aan bevindingen uit de praktijk, wordt ingezet op een kwalitatieve onderzoeksmethode. Omdat het hier het ontwikkelen van artefact betreft dat vervolgens getoetst wordt, zou design science (Peppers, K., Tuunanen, T., Rothenberger, M.A, Chatterjee, S., 2007) een alternatieve methode kunnen vormen. Vanwege het meer technisch georiënteerd toepassingsgebied van design science is hier niet voor gekozen.

##### *Doelstelling gerelateerd aan het ontwerp*

De doelstelling van een onderzoek kan meerdere vormen aannemen: verkennend, beschrijvend, verklarend, evaluerend of een combinatie van deze (Saunders, M.K.N, Lewis, P., Thornhill, A., 2019).

Dit onderzoek behelst het vaststellen van een lijst van termen en onderlinge relaties vanuit de literatuur, deze vervolgens toetsen aan de praktijk en resultaten hiervan, zowel vanuit een theoretisch als praktisch oogpunt, bij elkaar brengen ter beantwoording van de onderzoeksvraag. Bovenstaande heeft zowel een verkennende doelstelling met een inductieve benaderingswijze als een evaluerende/toetsende doelstelling met een deductieve benaderingswijze. Gegeven deze combinatie kan gesproken worden over een mixed method.

##### *Onderzoeksstrategie*

Voor een onderzoek zijn meerdere strategieën mogelijk (Saunders, M.K.N, Lewis, P., Thornhill, A., 2019). Het onderzoek heeft een kwalitatief karakter waarbij onderzocht wordt of termen en

onderlinge relaties, afgeleid vanuit de literatuur, herkend worden in de praktijk. Gegeven het bovenstaande worden 2 opties gezien: de termen en onderlinge relaties vertalen naar een enquête die bij meerdere organisaties wordt uitgezet of het uitvoeren van een casestudy met één of meerdere cases. Gegeven de tijd gemoeid met het achterhalen van geschikte organisaties, toetsen van de bereidwilligheid en deze vervolgens aanschrijven met een beperkte reactietermijn wordt een enquête, gezien de beschikbare tijdspanne en daarmee gepaard gaande haalbaarheid, als niet haalbaar geacht.

Een casestudy wordt daarmee als een geschiktere optie gezien waarbij wordt opgemerkt dat vanwege haalbaarheid deze wordt beperkt tot één case.

#### *Tijdshorizon*

Het onderzoek kan betrekking hebben op een specifieke tijdsperiode waarbij de stand van zaken op een bepaald moment als basis geldt (zgn. Cross-sectional studies) of een langere tijdsperiode omvatten waarbij veranderingen gedurende de tijd waargenomen kunnen worden (zgn. Longitudinal studies). Vanwege de beperkte tijdsspanne wordt gekozen van een cross-sectional benadering.

### 3.2 Technisch ontwerp: uitwerking van de methode

Het onderzoek valt uiteen in een theoretische benadering (bepalen van termen en onderlinge relaties o.b.v. de literatuur) en toetsing of de, vanuit de theorie onderkende, termen en relaties herkend worden in de praktijk middels het uitvoeren van een casestudy. Resultaten worden vervolgens bij elkaar gebracht ter beantwoording van de onderzoeksvraag.

De casestudy wordt uitgevoerd middels het houden van interviews. Andere alternatieven voor het uitvoeren van een casestudy zijn observaties en questionnaires. Een observatie is niet haalbaar omdat dan gedurende een lange tijd de transitie naar een CC oplossing gevolgd moet worden. Een questionnaire betekent dat je vooraf al een goed beeld moet hebben van het onderzoeksterrein, hetgeen niet het geval is.

Om herkenning van onderkende termen en onderlinge relaties te toetsen kunnen deze vertaald worden naar een aantal thema's/hoofdvragen die de leidraad vormen voor het interview.

Voor het houden van interviews kan gekozen worden tussen gestructureerde, semi-gestructureerde of ongestructureerde interviews waarbij qua samenstelling (één op één interview, groepsinterviews, workshops) of vorm (fysiek, telefoon, video) ook nog keuzes te maken zijn (Saunders, M.K.N, Lewis, P., Thornhill, A., 2019).

Gegeven het exploratieve karakter van het onderzoek, waarbij de mogelijkheid aanwezig moet zijn om vanuit de praktijkcasus nieuwe inzichten te verwerven, wordt een gestructureerd interview met questionnaires niet geschikt geacht. Een ongestructureerd interview waarbij alle vragen ontstaan tijdens het interview is ook niet van toepassing. De, vanuit de theorie onderkende, termen en onderlinge relaties kunnen immers vertaald worden naar een aantal richtinggevend vragen. Om daarbij tevens de ruimte te bieden aanvullende inzichten vanuit de praktijk te kunnen leveren, wordt gekozen voor een semi-gestructureerde invalshoek. Interviews worden gehouden in één op één sessies vanwege het kennisintensieve karakter van het onderzoeksonderwerp en daarbij gewenste interactie (kunnen volgen en reageren op antwoorden). Vanwege de COVID-19 situatie wordt, naar verwachting, voor een groot deel gebruik gemaakt van telefoon- en videoconferentie als vorm. Voor de in hoofdstuk 2 onderkende deelgebieden vinden interviews plaats met mensen uit verschillende disciplines zodat de problematiek vanuit verschillende invalshoeken bekeken kan worden. Onderstaand een overzicht van disciplines betrokken in het onderzoek en de relevante deelgebieden<sup>1</sup> waarover zij bevraagd worden.

---

<sup>1</sup> Medewerkers kunnen, uitgaande van hun rol, ook bevraagd over andere deelgebieden

Discipline	Deelgebieden			
	Samenhang Data Quality Management en Big Data Analytics	Data Quality Management uitgedrukt in de lagen van een Enterprise Architecture	Impact van Cloud Computing voor Big Data Analytics op de Enterprise Architecture	Enterprise Architecture Management als instrument voor impactbepaling bij invoering Cloud Computing
Data Steward	X	X	X	
Data Analyst	X	X		
Data Beheer		X		
Enterprise Architect		X	X	X
Informatie Architect		X	X	X
Teammanager				X

Figuur 4: Disciplines betrokken in het interview en deelgebieden waarover zij bevraagd worden

Resultaten uit de interviews worden teruggekoppeld richting de geïnterviewden middels gespreksverslagen met het verzoek te reageren en onvolledigheden/onjuistheden aan te geven waardoor eventuele misinterpretaties gesignaleerd kunnen worden. Geïnterviewden hebben daarbij de mogelijkheid aanvullingen en correcties aan te reiken.

Tijdens de interviews wordt tevens gevraagd of uitspraken kunnen worden onderbouwd middels het overhandigen van relevante achtergronddocumentatie, voor zover mogelijk.

### 3.3 Gegevensanalyse

Gegeven het kwalitatieve karakter van het onderzoek wordt een kwalitatieve gegevensanalyse toegepast. Hiervoor zijn verschillende technieken beschikbaar (Saunders, M.K.N, Lewis, P., Thornhill, A., 2019).

In het kader van dit onderzoek willen we o.b.v. de antwoorden op gestelde vragen inzicht krijgen of, vanuit de theorie verkregen, termen en onderlinge relaties herkend worden in de praktijk. Gegeven dit doel wordt gekozen voor een thematische analyse waarbij de volgende stappen worden doorlopen:

- Data eigen maken op basis van de interviewverslagen
- Data uit de interviewverslagen coderen waarbij de te hanteren codes zijn afgeleid uit de, vanuit de theorie onderkende, termen en onderlinge relaties.
- Op basis van de mate waarin data uit de interviewverslagen te koppelen is aan de gehanteerde codes, vaststellen of de termen en relaties (zoals onderkend vanuit de theorie) herkenning oproepen.

Thematische analyse vormt een systematische, toegankelijke en flexibele aanpak en is geschikt voor onderzoeken waarbij een theoretische uiteenzetting in de praktijk getoetst wordt. Door codes af te leiden van de termen en onderlinge relaties uit de literatuur wordt een relatie gelegd tussen de theoretische benadering en de toetsing in de praktijk. Een risico vormt dat de analyse teveel geleid wordt door de codes, gedestilleerd uit vooraf gedefinieerde termen en relaties, terwijl de interviews relevante inzichten kunnen leveren die niet hieraan te relateren zijn. Om dit risico te ondervangen wordt voorzien in de mogelijkheid aanvullende codes te definiëren. Daarnaast kan de gegevensanalyse leiden tot vervolginterviews voor nadere verdieping.

### 3.4 Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten

Onderstaand wordt ingegaan op de mate waarin de onderzoeksopzet voldoet aan validiteit, betrouwbaarheid en ethische aspecten en welke maatregelen ter waarborging worden geïntroduceerd.

#### *Reflectie op interne validiteit*

Interne validiteit betreft de mate waarin het onderzoek consistent is uitgevoerd en tot juiste conclusies leidt (Saunders, M.K.N, Lewis, P., Thornhill, A., 2019), (Runeson, P., Höst, M., 2009). Voor het waarborgen van de interne validiteit wordt gebruik gemaakt van triangulatie.

Triangulatie houdt in dat vanuit meerdere gegevensbronnen gewerkt wordt. Door het interviewen van meerdere medewerkers met verschillende disciplines en gebruik te maken van achtergronddocumentatie wordt getracht hier invulling aan te geven.

#### *Reflectie op externe validiteit*

Externe validiteit betreft de mate waarin de onderzoeksresultaten te generaliseren zijn (Saunders, M.K.N, Lewis, P., Thornhill, A., 2019), (Runeson, P., Höst, M., 2009). De onderzoeksopzet gaat uit van een, op literatuur gebaseerde, termen en beoogt deze te toetsen in de praktijk aan de hand van één casus bij één organisatie. Door uit te gaan van, op literatuur gebaseerde, termen wordt getracht een bepaalde mate van generaliseerbaarheid te bereiken. Het vervolg (één casus bij één organisatie) levert echter een risico m.b.t. externe validiteit. Ondanks dat getracht wordt de interviewresultaten te ontdoen van organisatie-specifieke elementen wordt verwacht dat de generaliseerbaarheid beperkt is.

#### *Reflectie op betrouwbaarheid*

Betrouwbaarheid betreft de mate waarin het onderzoek reproduceerbaar is (Saunders, M.K.N, Lewis, P., Thornhill, A., 2019), (Runeson, P., Höst, M., 2009). De onderzoeksopzet gaat uit van termen en relaties die vanuit de literatuur zijn verkregen en die vervolgens o.b.v. een casus getoetst wordt aan de praktijk. Literatuur is verkregen vanuit een bepaalde, tijdsgebonden, werkwijze (zie sectie 2.1) waarbij gehanteerde zoektermen, zoekmachine, filter- en sorteercriteria zijn vastgelegd. Relevante artikelen zijn daarbij gedownload en in mappen gearchiveerd. Middels deze werkwijze wordt getracht een bepaalde mate van betrouwbaarheid rondom de samenstelling van de aandachtspuntenlijst te garanderen.

De termen en relaties worden middels een casus getoetst aan de praktijk waarbij gebruik wordt gemaakt van semi-gestructureerde interviews. Tijdens de interviews worden vragen ingeleid en wordt getoetst of de vraag is begrepen. Interviewresultaten worden teruggekoppeld in de vorm van gespreksverslagen naar de geïnterviewden met het verzoek om een reactie op volledigheid/juistheid. Verslaglegging van interviews, toegepaste codering en daaruit voortvloeiende analyse resultaten worden vastgelegd. Gegeven deze werkwijze wordt getracht een bepaalde mate van betrouwbaarheid voor dit onderdeel te garanderen.

#### *Reflectie op ethische aspecten*

Ten aanzien van ethische aspecten worden de volgende maatregelen genomen:

- Casus organisatie dient expliciete goedkeuring te geven voor het onderzoek en de bijdragen van medewerkers hierin
- Bijdragen van de medewerkers wordt geanonimiseerd.
- Voor zover sprake is van vertrouwelijke informatie wordt gehandeld conform de vertrouwelijkheidseisen zoals gesteld vanuit de casusorganisatie.
- Onderzoeksopzet en uitvoering voldoet aan de richtlijnen zoals vermeld in de The Netherlands Code of Conduct for Academic Practice (Association of Universities in the Netherlands (VSNU), 2014).



## 4. Resultaten

### 4.1 Uitvoering

Het onderzoek valt uiteen in twee delen:

- Op basis van literatuur komen tot termen en onderlinge relaties en deze vertalen naar een lijst met aandachtspunten
- Toetsen van de termen met onderlinge relaties en de lijst met aandachtspunten aan de hand van een praktijkcasus

#### *Termen met onderlinge relaties en vertaling naar aandachtspunten vanuit een theoretisch oogpunt*

De, in sectie 1.4 verwoorde, deelvragen zijn vertaald naar 4 deelgebieden zoals weergegeven in sectie 1.6. Deze deelgebieden zijn, in aanvulling op de samenvattende resultaten weergegeven in sectie 2.3, vanuit de beschikbare literatuur nader verdiept en vertaald in een set van termen en hun onderlinge relaties. Dit vormde basis voor formulering van een vragenlijst per deelgebied die als leidraad gold voor de interviews (zie hiervoor bijlage II tot en met V). Daarnaast zijn de, vanuit de verdieping onderkende, termen en onderlinge relaties bij elkaar gebracht om te komen tot een lijst van aandachtspunten ter beantwoording van de onderzoeksvraag vanuit de theorie.

#### *Toetsing aan de hand van een praktijkcasus*

Het vervolg van het onderzoek was om per deelgebied de, vanuit een theoretisch oogpunt verkregen, termen en relaties te toetsen op herkenbaarheid in de praktijk en de bevindingen te relateren aan de lijst met aandachtspunten ter beantwoording van de onderzoeksvraag.

Als casusorganisatie is gekozen voor een middelgrote overheidsorganisatie. Deze organisatie is recent gestart met BDA voor beantwoording van beleids- en uitvoeringsvraagstukken waarbij tevens, op onderdelen, CC kan worden ingezet.

Voor het onderzoek zijn in april 2021, middels één op één gesprekken en gebruik makend van een video conference faciliteit, acht medewerkers van de casusorganisatie geïnterviewd. Alhoewel de rollen van deze medewerkers niet helemaal overeen kwamen met de rollen zoals gedefinieerd in sectie 3.2. is er toch een match te maken. Zo zijn er 2 medewerkers die een nauwe link hebben met de rol van Data Stewards en Data Analyst, 1 medewerker met een rol overeenkomend met de rol van Data Analyst en 2 medewerkers met een Data Beheer rol. Daarnaast zijn medewerkers met de rol van Enterprise Architect, Informatie Architect en Teammanager geïnterviewd.

Voorafgaand aan het interview is een interviewprotocol en een vragenlijst geformuleerd die als leidraad gold voor het interview. Daarnaast is een proefinterview gehouden. Uit dit proefinterview is naar voren gekomen om voorafgaand aan het interview een lijst toe te sturen met omschrijvingen van de onderkende deelgebieden en definities van frequent gebruikte termen.

Resultaten uit de interviews zijn verwoord in een gespreksverslag. De gespreksverslagen zijn binnen 3 dagen na het interview naar de geïnterviewden gestuurd met het verzoek dit, waar nodig, aan te vullen/te corrigeren en tevens aan te geven als het verslag overeenstemt met hetgeen besproken is. Elke respondent heeft gereageerd. Bij 4 respondenten zijn daarbij een aantal kleine correcties/aanvullingen aangereikt.

Gespreksverslagen zijn vervolgens onderworpen aan een thematische analyse per deelgebied. Hierbij zijn tekstfragmenten in een Excel spreadsheet handmatig gekoppeld aan termen of relaties tussen termen zoals deze vanuit het theoretisch deel zijn onderkend (zie hiervoor bijlage II tot en met V). Door een koppeling te leggen met termen en relaties vanuit de theorie is getracht te toetsen of het inzicht dat vanuit de theorie is verkregen, herkend wordt in de praktijk. Tijdens het koppelen van deze tekstfragmenten zijn ook aanvullende termen en relaties geïntroduceerd die niet vanuit het theoretisch inzicht initieel zijn meegegeven.

Bevindingen per deelgebied zijn gepresenteerd in de secties 4.2 t/m 4.5.

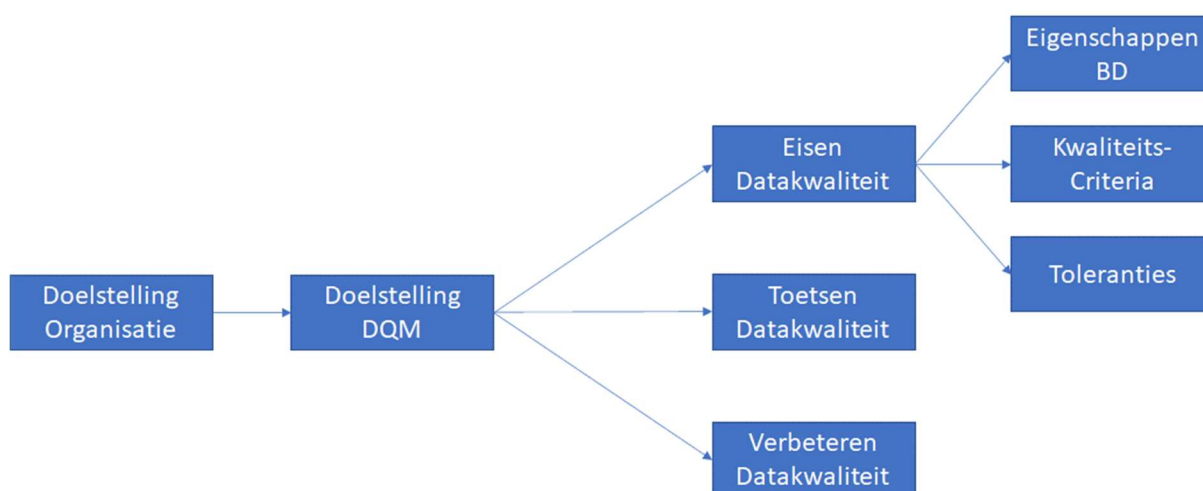
Bevindingen gekoppeld aan de hoofdvraag zijn weergegeven in sectie 4.6.



## 4.2 Op welke wijze hangt DQM samen met BDA?

### 4.2.1 Inzicht vanuit een theoretisch oogpunt

DQM is gericht op het borgen van de datakwaliteit middels het toetsen of data voldoet aan het vereiste kwaliteitsniveau en het uitvoeren van correctieve acties om dit niveau te bereiken. Het vereiste kwaliteitsniveau hangt daarbij samen met het doel wat de organisatie met de data voor ogen heeft en welke toleranties hierbij acceptabel zijn (Merino, J., Caballero, I., Rivas, B., Serrano, M., Piattini, M., 2015) (Ohbyung, K., Namyoon, L., Bongsik, S., 2014). Kwaliteitseisen kunnen uitgedrukt worden in een aantal criteria zoals compleetheid, consistentie etc. Big Data heeft echter een aantal specifieke eigenschappen vaak uitgedrukt in de 5 V's zijnde Volume, Velocity, Variety, Veracity en Value (Taleb, I., Serhani, M.A., Rchida, D., 2018). Voor kwaliteit van BD is het daarom zaak de kwaliteitseisen, uitgedrukt in kwaliteitscriteria en bijbehorende toleranties toe te passen op de eigenschappen die gelden voor BD (Merino, J., Caballero, I., Rivas, B., Serrano, M., Piattini, M., 2015). Onderstaande is een mogelijke visualisatie van het bovenstaande uitgedrukt in termen en hun onderlinge relaties.



Figuur 5: Samenhang DQM met BDA

#### Toelichting:

De doelstellingen voor DQM zijn een afgeleide van de doelstellingen voor de organisatie als geheel. DQM doelstellingen vertalen zich naar eisen voor datakwaliteit, maatregelen ter toetsing of aan deze eisen wordt beantwoord en maatregelen voor kwaliteitsverbetering. Eisen voor datakwaliteit zijn afhankelijk van de, van toepassing zijnde, kwaliteitscriteria, rekenschap houdende met de eigenschappen die gelden voor BD (de grondstof voor BDA) en toleranties die gehanteerd worden in het toepassen van de kwaliteitscriteria op BD.

### 4.2.2 Toetsing aan de praktijk

In de interviews is getoetst of de termen zoals genoemd in sectie 4.2.1 en hun onderlinge relaties herkend worden in de praktijk. Gestelde vragen, gegeven antwoorden en bijbehorende koppeling naar termen en onderlinge relaties is weergegeven in Bijlage II. In onderstaande tabel is aangegeven o.b.v. het koppelen van tekstfragmenten aan termen en/of relaties hoeveel keer een term en/of een relatie van termen gekoppeld is. Het aantal geeft daarmee een beeld van de mate waarin een term en/of relatie in de praktijk herkend wordt.

Term of relatie tussen termen	Aantal
Doelstelling Organisatie	3
Doelstelling Organisatie - Doelstelling DQM	2
Doelstelling Organisatie - Kwaliteitseisen	9
Doelstelling Organisatie - Doelstelling BDA	3
Doelstelling Organisatie - Toleranties	2
Doelstelling DQM	3
Doelstelling DQM - Kwaliteitseisen	1
Doelstelling BDA	4
Kwaliteitseisen	6
Kwaliteitseisen - Cloud Computing	2
Kwaliteitscriteria	3
Eigenschappen Big Data	6
Eigenschappen Big Data - Kwaliteitseisen	3
Eigenschappen Big Data - Toetsen Datakwaliteit	5
Eigenschappen Big Data - Cloud Computing	1
Toleranties	2
Toetsen Datakwaliteit	12
Toetsen Datakwaliteit - Cloud Computing	1
Verbeteren Datakwaliteit	5
Verantwoordelijkheden	2
Expertise BDA	2

Figuur 6: Gebruikte termen en relaties bij samenhang DQM met BDA

Kijkend naar het aantal keer dat een term aan een tekstfragment gekoppeld kan worden, valt op te maken dat elke term uit het schema in sectie 4.2.1 wel herkenning oproept.

Naast de termen zijn er ook relaties tussen termen (in het schema van sectie 4.2.1 aangegeven middels een pijltje, en in bovenstaande tabel aangegeven middels een liggend streepje). Als we kijken naar de tabel worden relaties tussen de termen, zoals onderkend in het schema van sectie 4.2.1, niet allemaal herkend.

Naast het tellen van de gebruikte termen is in onderstaande tabel per term, o.b.v. de interviewresultaten, een samenvattende inhoudelijke motivatie gegeven die de herkenning van een term ondersteunt. Naast termen vanuit de theorie, zijn termen benoemd die aanvullend vanuit het interview zijn aangedragen en in de context van het deelgebied relevant zijn.

Term	Oorsprong Term	Herkenning?	Motivatie vanuit praktijkcasus
Doelstelling Organisatie	Theorie	Ja	Doelstellingen van de organisatie is om op een effectieve en efficiënte manier maatschappelijke vraagstukken op te lossen. Hieraan liggen beslissingen ten grondslag waarvan de juistheid afhankelijk is van de onderliggende data
Doelstelling DQM	Theorie	Ja	Doelstellingen van DQM is te voldoen aan de kwaliteitseisen die vanuit het beleid gesteld worden zodat de juiste beslissingen genomen worden ter beantwoording van maatschappelijke vraagstukken
Eisen datakwaliteit	Theorie	Ja	De van toepassing zijnde kwaliteitseisen zijn afhankelijkheid van het vraagstuk wat speelt en moeten vanuit het beleid aangedragen worden
Kwaliteitscriteria	Theorie	Ja	Data kan vanuit verschillende kwaliteitscriteria belicht worden (bijvoorbeeld volledigheid, tijdigheid, integriteit etc.). Een kwaliteitseis is een invulling van een kwaliteitscriterium (bijvoorbeeld de data moet voor 95% volledig zijn). Deze invulling kan worden weergegeven in een GegevensLeverings-Protocol die als meetlat kan gelden voor het toetsen van de datakwaliteit
Eigenschappen BD	Theorie	Ja	BD heeft specifieke kenmerken (uitgedrukt in de 5 V's). Kwaliteitseisen die je stelt aan BD kunnen echter tot uitdrukking worden gebracht in kwaliteitscriteria die ook gelden voor andersoortige data, niet zijnde BD
Toleranties	Theorie	Ja	De van toepassing zijnde toleranties voor datakwaliteit zijn afhankelijk van de impact die een beslissing heeft en moeten vanuit het beleid aangedragen worden
Toetsen datakwaliteit	Theorie	Ja	Voor het toetsen van de datakwaliteit van BD moet getracht worden de data om te zetten naar een zodanige vorm dat hier reeds bekende, vanuit gestructureerde data beschikbare, technieken kunnen worden ingezet. Het toetsen van de datakwaliteit wil je zo veel mogelijk reeds beleggen bij de bronhouder (eigenaar) van de data
Verbeteren datakwaliteit	Theorie	Ja	Verbeteracties voor datakwaliteit moeten belegd worden bij degene die als bronhouder verantwoordelijk is voor de data. Organisatie werkt, naast interne bronhouders, veel met externe bronhouders
Doelstelling BDA	Interview	Aanvullend	BDA kan helpen met het oplossen van vraagstukken maar is geen doel op zich. Doel moet voortkomen uit de beleidsvraagstukken

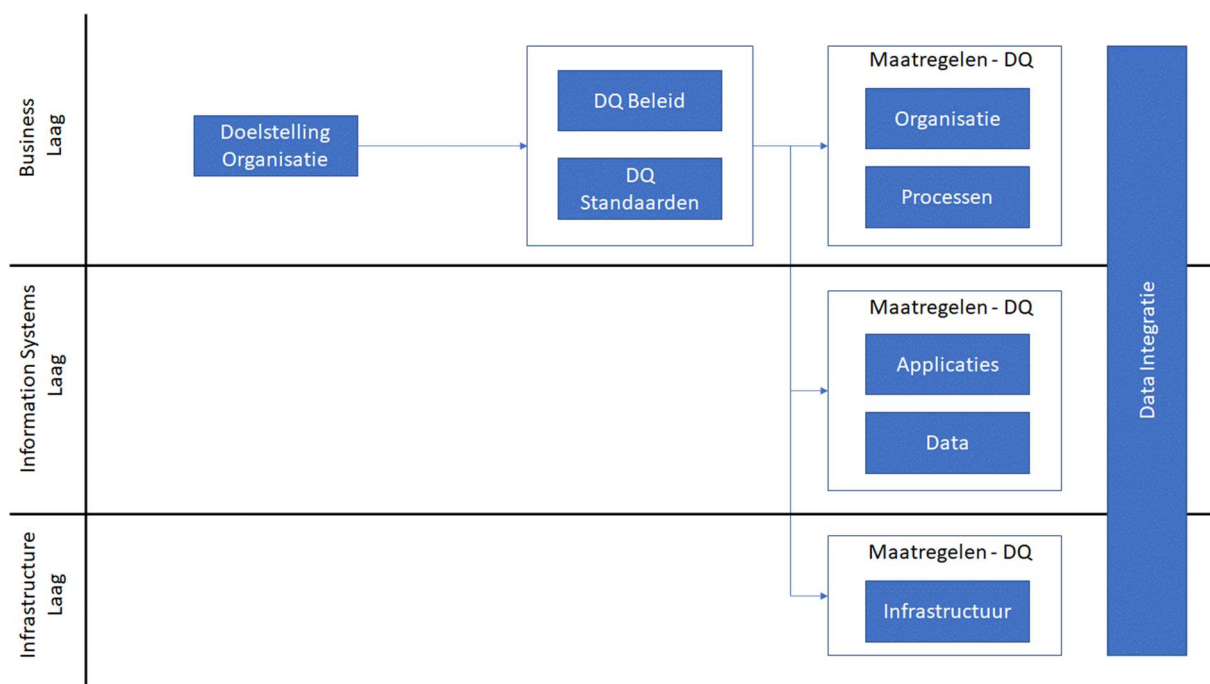
Figuur 7: Onderbouwing herkenning van termen bij samenhang tussen DQM en BDA

## 4.3 Hoe manifesteert DQM zich in de lagen van een EA?

### 4.3.1 Inzicht vanuit een theoretisch oogpunt

DQM manifesteert zich in alle lagen van een EA (Panian, Z., 2016) (Kim, H.Y., Cho, J-S., 2017) (Burmeister, F., Drews, P., Schirmer, I., 2018). Op het niveau van de Business laag komt dit tot uitdrukking in beleid en standaarden voor datakwaliteit, afgeleid uit de doelstellingen die organisatie met de data voor ogen heeft en een daaraan gekoppeld kwaliteitsniveau. Beleid en standaarden vertalen zich vervolgens naar een procesinrichting waarbij aandacht is voor (een continu proces van) toetsing van de datakwaliteit en corrigerende acties om de datakwaliteit te verbeteren. Daarnaast moet op het niveau van de Business laag aandacht worden besteed aan de organisatiecomponent uitgedrukt in de rollen en verantwoordelijkheden voor het borgen van de datakwaliteit en daarvoor benodigde expertises. Op het niveau van de Information Systems laag vertaalt zich dit naar toepassingen waarin aandacht is voor het (geautomatiseerd) afdwingen van datakwaliteit, het detecteren van datakwaliteit-issues en het bieden van voorzieningen om data te corrigeren/schonen. Beleid en de standaarden voor datakwaliteit komen ook tot uitdrukking in de wijze van opslag van data, wat zowel de Information Systems laag als de Infrastructurele laag raakt. Een voornaam aandachtspunt dat alle lagen van een EA raakt, betreft de mate van data-integratie waarbij afgestapt wordt van afzonderlijke, geïsoleerde databronnen en toegewerkt wordt naar een organisatie-brede, geïntegreerde gegevensverzamelingen (Panian, Z., 2016) (Comuzzi, M., Patel, A., 2016).

Onderstaande is een mogelijke visualisatie van het bovenstaande uitgedrukt in termen en hun onderlinge relaties.



Figuur 8: DQM uitgedrukt in een EA

#### Toelichting:

De EA van een organisatie kan vertaald worden naar 3 lagen: de Business laag, de Information Systems laag en de Infrastructurele laag. De doelstellingen van de organisatie op het gebied van datakwaliteit leiden tot beleid en standaarden voor datakwaliteit (DQ). Beleid en standaarden vertalen zich op hun beurt naar maatregelen voor toetsing en verbetering van datakwaliteit op

niveau van de Business laag (organisatie en processen), Information Systems laag (applicaties en data) en de Infrastructuur. Een belangrijk aspect vormt hierbij de data-integratie die tot uitdrukking komt in alle lagen van de EA.

#### 4.3.2 Toetsing aan de praktijk

In de interviews is getoetst of de termen zoals genoemd in sectie 4.3.1 en hun onderlinge relaties herkend worden in de praktijk. Gestelde vragen, gegeven antwoorden en bijbehorende koppeling naar termen en onderlinge relaties is weergegeven in Bijlage III. In onderstaande tabel is aangegeven o.b.v. het koppelen van tekstfragmenten aan termen en/of relaties hoeveel keer een term en/of een relatie van termen gekoppeld is. Het aantal geeft daarmee een beeld in de mate waarin een term en/of relatie vanuit de praktijk herkend wordt.

Term of relatie tussen termen	Aantal
Adoptie EA	5
Adoptie EA voor DQM	6
Adoptie EAM	6
Doelstelling Organisatie - Datakwaliteit	2
Doelstelling Organisatie - Datakwaliteit - Inrichting Toetsen	1
Doelstelling Organisatie - Datagebruik	1
Doelstelling DQM	1
Doelstelling DQM - Beleid en Standaarden voor DQ	1
Beleid en Standaarden voor DQ	15
Standaarden Algemeen	1
Processen	5
Processen - IT systemen	1
Verantwoordelijkheden	7
Verbeteren Datakwaliteit - Verantwoordelijkheden	1
IT Systemen	5
Kwaliteitseisen	1
Big Data - Datakwaliteit - Inrichting Toetsen	3
Maatregelen Borgen Datakwaliteit	4
Datakwaliteit - Inrichting Toetsen	3
Toetsen Datakwaliteit	1
Data Beheer	7
Datagebruik	9
Verbetering Inrichting	2
Ketensamenwerking	1
Integratie	1

*Figuur 9: Gebruikte termen en relaties bij DQM uitgedrukt in een EA*

Kijkend naar het aantal keer dat een term aan een tekstfragment gekoppeld kan worden, valt op te maken dat bijna elke term uit het schema in sectie 4.3.1 herkenning oproept.

Er zijn aanvullende termen te onderkennen waarbij een aantal termen niet rechtstreeks met het deelgebied van doen hebben maar in de context van het deelgebied wel relevant zijn (bijv. adoptie EAM en EA).

Daarnaast zijn een aantal termen benoemd die niet meteen rechtstreeks te koppelen zijn aan de, vanuit de theorie onderkende, termen maar er wel een relatie naar te maken is. Zo is verantwoordelijkheden te koppelen aan organisatie, databeheer/gebruik te koppelen aan data en ketensamenwerking aan integratie.

Naast de termen zijn er ook relaties tussen termen (in het schema van sectie 4.3.1. aangegeven middels een pijltje en in bovenstaande tabel aangegeven middels een liggend streepje). Als we kijken naar de tabel kunnen relaties zoals onderkend in het schema van sectie 4.3.1. niet allemaal herkend worden.

Naast het tellen van de gebruikte termen is in onderstaande tabel per term, o.b.v. de interviewresultaten, een samenvattende inhoudelijke motivatie gegeven die de herkenning van een term ondersteunt. Naast termen vanuit de theorie, zijn termen benoemd die aanvullend vanuit het interview zijn aangedragen en in de context van het deelgebied relevant zijn.



Term	Oorsprong Term	Herkenning?	Motivatie vanuit praktijkcasus
Doelstelling Organisatie	Theorie	Ja	Afhankelijk van het doel wat de organisatie met de data voor ogen heeft worden eisen ten aanzien van de datakwaliteit geformuleerd
Beleid en standaarden DQ	Theorie	Ja	Beleid en standaarden voor datakwaliteit moeten top-down en eenduidig voorgeschreven worden. Momenteel is sprake van een aantal basisprincipes (zoals enkelvoudig beheer een meervoudig gebruik) en een aantal standaarden gericht op metadata maar dit geheel kan meer uitgebreid worden
Maatregelen DQ voor processen en organisatie	Theorie	Ja	Maatregelen voor datakwaliteit moeten, alhoewel hier voor de organisatie nog verbeteringen in mogelijk zijn, vertaald worden naar een structurele en formele inbedding in processen en bijbehorende organisatie uitgedrukt in rollen/verantwoordelijkheden, rekening houdend met het onderscheid in gegevensbeheer en gegevensgebruik
Maatregelen DQ voor applicaties en data	Theorie	Ja	Belang van het (geautomatiseerd) toetsen en afdwingen van datakwaliteit neemt toe, zeker gezien de toename van de omvang van data en het belang van BD. Daarbij is het van belang dat goed onderscheid worden gemaakt in databeheer en datagebruik en het beheer van data eenduidig belegd hebt (dus niet meerdere systemen voor hetzelfde doel)
Maatregelen DQ voor Integratie	Theorie	Niet behandeld	Niet aan bod gekomen tijdens interviews
Adoptie EAM/EA algemeen	Interview	Aanvullend	Data-integratie is van belang waarbij afgestapt wordt van het in silo's denken. Daarbij gaat deze integratie verder dan alleen de interne organisatie. Gezien het aantal externe bronhouders wordt ketensamenwerking steeds belangrijker
Adoptie EA voor DQM	Interview	Aanvullend	Organisatie hanteert een EA afgeleid van een overheids-brede referentiearchitectuur waarop een actualisatieslag gaande is. Naast een EA maakt de organisatie per project gebruik van een Project Start Architectuur waarin voor desbetreffende project wordt onderbouwd in welke mate gebruik wordt gemaakt van geldende architectuurprincipes
			Enerzijds wordt door de organisatie het nut van een EA voor DQM onderkend om risico's en maatregelen omtrent datakwaliteit in beeld te brengen. Anderzijds wordt aangegeven dat een EA mogelijk te hoog over is en onderdelen voor datakwaliteit voornamelijk in de Information Systems laag landen

Figuur 10: Onderbouwing herkenning van termen voor DQM uitgedrukt in een EA

## 4.4 Op welke wijze grijpt CC voor BDA in op een EA?

### 4.4.1 Inzicht vanuit een theoretisch oogpunt

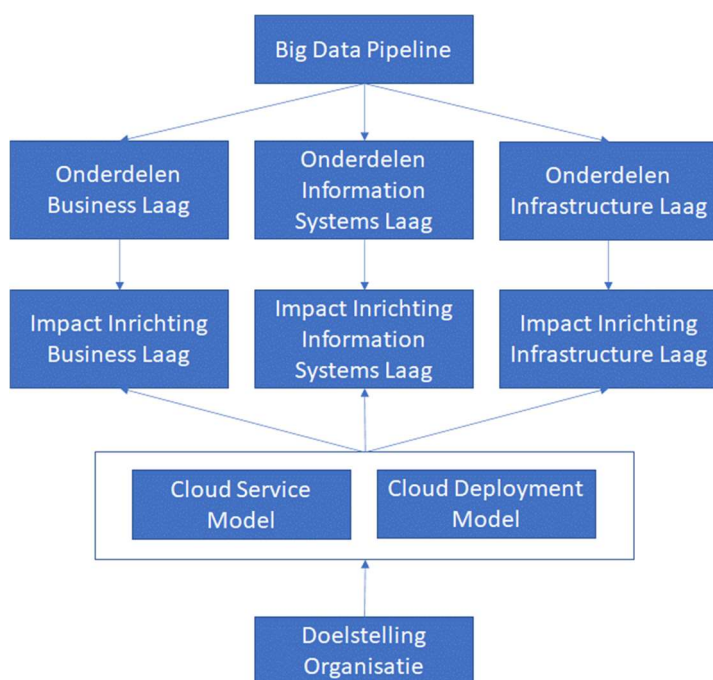
Het proces om vanuit de databronnen voor Big Data te komen tot data-analyse resultaten wordt aangeduid met de term Big Data Pipeline (Lnenicka, M., Komarkova, J., 2018). De stappen die onderdeel vormen van deze Big Data Pipeline kunnen ingericht worden op de verschillende lagen van een EA. De wijze waarop dit ondersteund kan worden middels CC is afhankelijk van het gekozen Cloud Service Model en Cloud Deployment model.

Het Cloud Service Model duidt het type services wat vanuit de Cloud beschikbaar wordt gesteld (Mahmood, Z., Hill, R., 2011). Gangbare typeringen zijn hierbij Infrastructure as a Service (IaaS), Platform as a Service (PaaS) en Software as a Service (SaaS).

Het Cloud Deployment Model duidt de wijze waarop Cloud services beschikbaar worden gesteld. Hierbij wordt onderscheid gemaakt in een Public Cloud waarbij Cloud Services verstrekt worden vanuit (commerciële) leveranciers, een Private Cloud waarbij de Cloud services verstrekt worden vanuit datacenters van de organisatie zelf en een Hybrid Cloud, zijnde een combinatie van de twee eerdere mogelijkheden.

De keuze van het te gebruiken Cloud Service Model en Cloud Deployment Model is afhankelijk van de doelstellingen die de organisatie nastreeft (Schmidt, R., Möhring, M., 2013). Afhankelijk van deze keuze grijpt CC in op de verschillende lagen van een Enterprise Architecture (Mahmood, Z., Hill, R., 2011).

Onderstaande is een mogelijke visualisatie van het bovenstaande uitgedrukt in termen en hun onderlinge relaties.



Figuur 11: Wijze waarop CC voor BDA ingrijpt op een EA

#### Toelichting:

BDA komt tot uitdrukking in stappen van een Big Data Pipeline waarvan de inrichting zich vertaalt in onderdelen van een Business laag, een Information Systems laag en/of Infrastructurele laag. Doelstellingen van de organisatie vertalen zich naar de keuze voor het te hanteren Cloud Service Model en Cloud Deployment Model. Afhankelijk van de keuze voor het te hanteren Cloud Service en Cloud Deployment Model heeft dit impact op de ingerichte onderdelen van de Big Data Pipeline voor de Business, Information Systems en Infrastructurele laag.

#### 4.4.2 Toetsing aan de praktijk

In de interviews is getoetst of de termen zoals genoemd in sectie 4.4.1 en hun onderlinge relaties herkend worden in de praktijk. Gestelde vragen, gegeven antwoorden en bijbehorende koppeling naar termen en onderlinge relaties is weergegeven in Bijlage IV. In onderstaande tabel is aangegeven o.b.v. het koppelen van tekstfragmenten aan termen en/of relaties hoeveel keer een term en/of een relatie van termen gekoppeld is. Het aantal geeft daarmee een beeld in de mate waarin een term en/of relatie vanuit de praktijk herkend wordt.

Term of relatie tussen termen	Aantal
Big Data Pipeline	2
Doelstelling Organisatie - Doelstelling CC	8
Doelstelling Organisatie - Toe te passen Cloud Service en Cloud Deployment Model	2
Doelstelling CC	1
Cloud Service Model - Effect op EA lagen - Datakwaliteit	2
Cloud Deployment Model - Effect op EA lagen - Datakwaliteit	1
Datakwaliteit	2
DataKwaliteit - Inzet CC	1
Cloud Deployment Model	1
Data Integratie	1

Figuur 12: Gebruikte termen en relaties m.b.t. wijze waarop CC voor BDA ingrijpt op een EA

Het interviewresultaat voor dit deelgebied is beperkt, gegeven de beginfase waarin de organisatie zich bevindt in het toepassen van BDA en de ondersteuning van CC hierin.

Kijkend naar het aantal keer dat een term aan een tekstfragment gekoppeld kan worden, valt op te maken dat termen zoals onderkend vanuit een theoretisch oogpunt niet allemaal expliciet herkend worden in het interview. Hetzelfde geldt m.b.t. de relaties tussen de termen.

Naast het tellen van de gebruikte termen is in onderstaande tabel per term, o.b.v. de interviewresultaten, een inhoudelijke motivatie gegeven die de herkenning van een term ondersteunt. Naast termen vanuit de theorie, zijn termen benoemd die aanvullend vanuit het interview zijn aangedragen en in de context van het deelgebied relevant zijn.

Term	Oorsprong Term	Herkenning?	Motivatie vanuit praktijkcasus
Big Data PipeLine	Theorie	Ja, maar niet uniform	Organisatie onderkend een Big Data Pipeline. Respondenten leggen deze echter verschillend uit als deze vertaald wordt naar de stappen die onderdeel vormen van de Big Data Pipeline
Onderdelen Business, Information Systems en Infrastructure laag	Theorie	Deels	Behalve het benoemen van de stappen die onderdeel vormen van een Big Data Pipeline is dit in de interviews niet nader tot uitdrukking gekomen
Doelstelling Organisatie	Theorie	Ja	Wijze waarop je CC gaat inzetten moet zijn grondslag hebben in de doelstelling van de organisatie waarbij sprake is van een top-down benadering
Cloud Service en Deployment Model	Theorie	Ja	Vanuit de doelstelling van de organisatie moet toegewerkt worden naar het te hanteren Cloud Service en Cloud Deployment Model. Organisatie ziet dat dit nog een nadere invulling moet krijgen. Organisatie heeft wel als richtlijn dat bij introductie van nieuwe toepassingen kiest voor SaaS oplossingen. Invloed van de markt moet daarbij niet onderschat worden. On premise oplossingen worden bijna niet meer geboden. Toegespitst op BDA is de toepassing van CC nog incident gedreven bij het ondervinden van knelpunten in de infrastructuur. Dit moet echter ook gezien worden in het kader van de beginfase waarin de organisatie zich bevindt in et toepassen van BDA.
Impact CC op inrichting Business, Information Systems en Infrastructure laag	Theorie	Deels	Afhankelijk van de keuze voor het te hanteren Cloud Service Model heeft dit consequenties voor de invulling van de Business, Information Systems en Infrastructure laag. Deze lagen zijn echter niet expliciet in het interview naar voren gekomen. Wel is geconstateerd dat hoe meer je echter opschuift naar de Business laag hoe minder controle je als organisatie zelf kunt uitoefenen. Ten aanzien van de keuze voor het te hanteren Cloud Deployment Model wordt aangegeven dat bij een private Cloud je alles in eigen hand houdt terwijl bij een public Cloud dit uit handen geeft. Door te kiezen voor een hybride vorm beperkt je deze afhankelijkheid enigszins
Datakwaliteit	Interview	Aanvullend	Het borgen van de datakwaliteit en bijbehorende data-integratie moet een onderdeel van alle stappen binnen een Big Data Pipeline. De keuze voor het Cloud Service en Deployment Model heeft effect op de wijze waarop de datakwaliteit geborgd is en in hoeverre je dit nog in eigen handen hebt
Dataintegratie	Interview	Aanvullend	Dataintegratie vormt een belangrijke factor bij het inrichten van de Big Data Pipeline

Figuur 13: Onderbouwing herkenning van termen bij wijze waarop CC voor BDA ingrijpt op een EA.

## 4.5 Hoe kan EAM ingezet worden voor impactbepaling bij migratie naar CC?

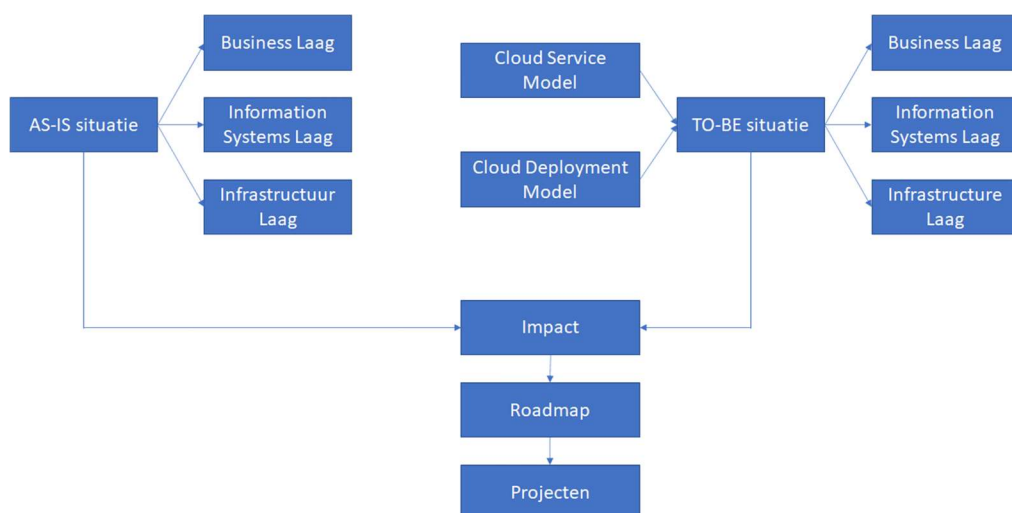
### 4.5.1 Inzicht vanuit een theoretisch oogpunt

EAM is een methode voor het beheerst doorvoeren van de veranderingen in de inrichting van organisaties waarbij de huidige en gewenste inrichting zijn uitgedrukt in de lagen van een EA (Ahleman F., Stettiner E., Messerschmidt, M., Legner, C., 2012).

Een EA kan een hulpmiddel vormen bij migratie naar CC waarbij, naast het voorzien in een EA, aandacht moet worden besteed aan de samenhang tussen de componenten waarbij onnodige verwevenheid moet worden vermeden en de wijze waarop deze componenten aansluiten op het Cloud Service Model en Cloud Deployment Model dat de organisatie voor ogen heeft (Mahmood, Z., Hill, R., 2011).

Door de AS-IS situatie en de TO-BE situatie te vergelijken kan de impact bepaald worden die zich vervolgens vertaalt naar een Roadmap met uit te voeren projecten (Ahleman F., Stettiner E., Messerschmidt, M., Legner, C., 2012).

Bovenstaande theoretische uiteenzetting is vertaald worden naar onderstaande visualisatie.



Figuur 14: EAM als instrument voor impactbepaling bij migratie naar CC

#### Toelichting:

Om impact kunnen bepalen is het van belang inzicht te hebben in de huidige (AS-IS) situatie en de gewenste (TO-BE) situatie, beide uitgedrukt in de lagen van een EA, zijnde de Business, Information Systems en Infrastructuur laag. De gewenste situatie is een, middels CC ondersteunde, situatie waarbij keuzes in het te hanteren Cloud Service Model en Cloud Deployment Model vertaald zijn in (aangepaste) onderdelen van de 3 lagen van een EA.

De impact vormt de basis voor een Roadmap met uit te voeren projecten om vanuit de huidige situatie te migreren naar de gewenste situatie.

#### 4.5.2 Toetsing aan de praktijk

In de interviews is getoetst of de termen zoals genoemd in sectie 4.5.1 en hun onderlinge relaties herkend worden in de praktijk. Gestelde vragen, gegeven antwoorden en bijbehorende koppeling naar termen en onderlinge relaties is weergegeven in Bijlage V. In onderstaande tabel is aangegeven o.b.v. het koppelen van tekstfragmenten aan termen en/of relaties hoeveel keer een term en/of een relatie van termen gekoppeld is. Het aantal geeft daarmee een beeld in de mate waarin een term en/of relatie vanuit de praktijk herkend wordt

Term of relatie tussen termen	Aantal
Adoptie EAM Algemeen	6
Adoptie EAM - Veranderkundig	2
Adoptie EAM - Impact - Veranderkundige aspecten	2
Adoptie EAM - Roadmap - Veranderkundige aspecten	2
Adoptie EAM - Monitoring	4
Adoptie EA Algemeen	6
Adoptie EA - DQM	2
Adoptie EA - BDA	1
Adoptie EA - Weergave ASIS	4
Adoptie EA - Weergave TOBE	1
Projecten - Project Start Architectuur	2
Projecten - Project Start Architectuur - DQM	1
Projecten - Project Start Architectuur - Programma van Eisen	1
Projecten - ProjectBrief	1
Projecten - ProjectBrief - Goedkeuring	1
Projecten - ProjectBrief - Project Start Architectuur/Projectplan	1
Doelstelling Organisatie - Doelstelling CC	9
Eisen CC	1
Toepassing CC	3
Toepassing CC - TOBE - EA Lagen	4
Toepassing CC - Leveranciers	2
Toepassing CC - Leveranciers - Infoveiliging & Beheer	1

Figuur 15: Gebruik termen en relaties m.b.t. inzet EAM voor impactbepaling bij migratie naar CC



Uit bovenstaande tabel valt op te maken dat, alhoewel de AS-IS en TO-BE situatie vermeld worden, niet alle termen in de interviews naar voren zijn gekomen. Hetzelfde geldt voor de relaties tussen de termen. Een groot deel gaat over adoptie van EAM als methodiek en het gebruik van EA als middel voor weergave van een AS-IS en TO-BE situatie in zijn algemeenheid en voor BDA/DQM in het bijzonder. Daarnaast is de toepassing van CC, de vertaling naar de TO-BE situatie en rol van leveranciers aan bod gekomen.

Naast het tellen van de gebruikte termen is in onderstaande tabel per term, o.b.v. de interviewresultaten, een samenvattende inhoudelijke motivatie gegeven die de herkenning van een term ondersteunt. Naast termen vanuit de theorie, zijn termen benoemd die aanvullend vanuit het interview zijn aangedragen en in de context van het deelgebied relevant zijn.

Term	Oorsprong Term	Herkenning?	Motivatie vanuit praktijkcasus
AS-IS situatie, uitgedrukt in Business, Information Systems en Infrastructure laag	Theorie	Deels	Weergave van een IST situatie is essentieel. Dit vormt het vertrekpunt om impact te kunnen bepalen. Als je ergens naar toe wilt moet je weten waar je vandaan komt en dat is meestal geen groene weide. In interviews is dit echter niet verder uigesplitst naar de diverse lagen van een EA.
Cloud Service en Deployment Model	Theorie	Ja	Organisatie onderkent het nut van de vertaling van organisatiedoelstellingen naar een Cloud Service en Cloud Deployment model. Wijze van ondersteuning middels CC vormt onderdeel van een reeds geformuleerde Sourcingsstrategie die nog een andere invulling moet krijgen. Marktpartijen zetten reeds volop in op CC oplossingen en als organisatie dien je hier rekening mee te houden
TO-BE situatie, uitgedrukt in Business, Information Systems en Infrastructure laag	Theorie	Deels	Komen tot een TO-BE situatie is een organisatie-breed vraagstuk waarbij o.b.v. eisen/randvoorwaarden een duurzame oplossing wordt geschetst. In de totstandkoming van deze TO-BE situatie dient niet alleen aandacht te worden besteed aan datgene waar de organisatie invulling aan geeft maar ook aan datgene waaraan zij geen invulling geeft, in lijn met de organisatiedoelstellingen. Bij de vertaling van een Cloud Service en Deployment Model naar een TO-BE situatie kunnen in eerste instantie meerdere varianten uitgewerkt worden die basis vormen voor nadere besluitvorming. In interviews is dit echter niet verder uigesplitst naar de diverse lagen van een EA
Impact	Theorie	Ja	Weergave van een AS-IS en TO-BE situatie geeft houvast bij het uitvoeren van veranderingstrajecten en zorgt voor bewustzijn over de consequenties gekoppeld aan een migratietraject. Belangrijk is om hierbij aandacht te besteden aan de zachte (veranderkundige) kant van een migratietraject. Veranderingen vormen immers een onzekere factor en je moet het vertrouwen van de medewerkers winnen
Roadmap	Theorie	Ja	Impact moet vertaald worden in uit voeren projecten. Daarbij is van belang in deze roadmap rekening te houden met het verandervermogen van de organisatie. Het is in dit kader vaak beter te voorzien in geleidelijke migraties waarbij kleine resultaten worden geboekt die uiteindelijk bijdragen in het einddoel
Projecten	Theorie	Ja	Bij de initiatie van projecten wordt gewerkt met een projectbrief die ter goedkeuring worden voorgelegd, gevolg door het opstellen van een Project Start Architectuur (PSA). De PSA creëert bewustzijn ten aanzien van allerlei aspecten die spelen bij een implementatie/migratie. Bij projectuitvoering is monitoring van de projecten essentieel die enerzijds kan leiden tot bijstelling van de projecten maar anderzijds kan leiden tot bijstellen van de TO-BE architectuur
Adoptie EAM en EA	Aanvullend	Aanvullend	Alhoewel het nut van een EAM en een EA wordt onderkend is het gebruik van een EAM en EA bij het doorvoeren van veranderingen niet bij alle respondenten bekend. Daar bovenop wordt het nut van toepassing van een EA voor BDA en DQM door een aantal respondenten betwijfeld daar men verwacht dat het merendeel van de componenten gaat landen in de Information Systems Laag

Figuur 16: Onderbouwing herkenning van termen voor inzet EAM voor impactbepaling bij migratie naar CC

## 4.6 EAM als instrument voor impactbepaling op DQM bij CC voor BDA

### 4.6.1 Synthese vanuit een theoretisch oogpunt

De hoofdvraag heeft betrekking op de wijze waarop EAM kan worden ingezet bij het bepalen van de impact voor DQM wanneer BDA migreert naar een CC oplossing met focus op de Business laag van een EA. Beantwoording van deze vraag vertaalt zich naar een set van aandachtspunten. M.a.w.: voor zinvolle inzet van EAM moet rekening worden gehouden met de benoemde aandachtspunten. Kijkend naar de context van de hoofdvraag kunnen de, in onderstaande tabel weergegeven, relevante termen per deelgebied onderkend worden.

Deelgebied	Termen	Onafhankelijk	AS-IS	TO-BE
Samenhang DQM - EA	Doelstelling Organisatie	X		
	Doelstelling DQM	X		
	Eisen voor Datakwaliteit (DQ)	X		
	Eigenschappen Big Data	X		
	Kwaliteitscriteria	X		
	Toleranties	X		
	Toetsing van Datakwaliteit		X	X
	Verbetering van Data Kwaliteit		X	X
DQM uitgedrukt in een EA	Doelstelling Organisatie	X		
	DQ Beleid en Standaarden	X		
	Maatregelen DQ op niveau van de Organisatie		X	X
	Maatregelen DQ op niveau van de Processen		X	X
Wijze waarop CC voor BDA ingegrijpt op een EA	Big Data Pipeline		X	X
	Onderdelen Business Laag		X	X
	Doelstelling Organisatie	X		
	Cloud Service Model			X
	Cloud Deployment Model			X
	Impact Inrichting Business Laag		X	X
Inzet EAM voor impactbepaling bij migratie naar CC	AS-IS situatie		X	
	Business Laag		X	
	Cloud Service Model			X
	Cloud Deployment Model			X
	TO BE situatie			X
	Business Laag			X
	Impact		X	X

Figuur 17: Relevante termen per deelgebied ter beantwoording hoofdvraag

#### Toelichting tabel:

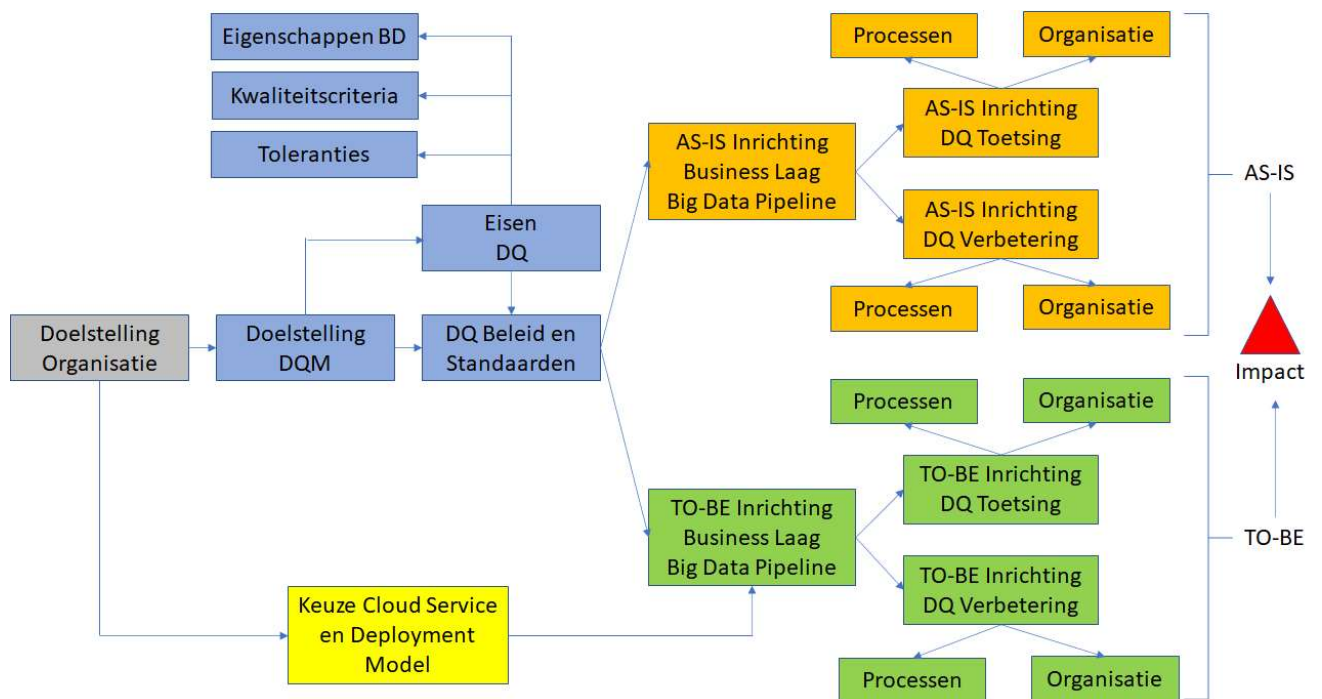
Per deelgebied is nagegaan welke termen relevant zijn kijkend naar de hoofdvraag.

Hierbij zijn (deels overlappende) termen te onderkennen met onderscheid in termen die onafhankelijk zijn van de AS-IS of TO-BE inrichting (zgn. inrichtingsonafhankelijke termen) en termen die wel afhankelijk zijn van de AS-IS of TO-BE inrichting (zgn. inrichtingsafhankelijke termen). Dit onderscheid is weergegeven middels de kruisjes in de drie laatste kolommen.

Vervolgens zijn de termen middels een kleur gecategoriseerd:

- Wat betreft de inrichtingsonafhankelijk termen:
  - Naar termen gericht op Doelstelling van de Organisatie (grijs)
  - Naar termen gericht op Datakwaliteit (blauw)
- Wat betreft de Inrichtingsafhankelijke termen:
  - Naar termen gericht op de AS-IS situatie voor Toetsing en Verbetering van Datakwaliteit (oranje)
  - Naar termen gericht op de TO-BE situatie voor Toetsing en Verbetering van datakwaliteit (groen)
  - Naar termen gericht op keuze voor het Cloud Service Model en Cloud Deployment Model (geel)

Met bovenstaande verdeling en rekening houdend met de relatie tussen termen vanuit het theoretisch inzicht voor de afzonderlijke deelgebieden kan onderstaand beeld geconstrueerd worden.



Figuur 18: Samenhang termen ter beantwoording hoofdvraag

#### Toelichting schema:

Organisatiedoelstellingen vertalen zich naar doelstellingen voor DQM die op hun beurt vertaald kunnen worden naar eisen voor datakwaliteit en bijbehorend beleid en standaarden. Eisen voor datakwaliteit worden uitgedrukt in de toegepaste kwaliteitscriteria, rekening houdend met de eigenschappen voor BD, en bijbehorende toleranties.

Beleid en standaarden voor datakwaliteit vertalen zich op niveau van de Business laag van een EA naar een proces en -organisatie-inrichting van de Big Data Pipeline met aandacht voor toetsing en verbetering van datakwaliteit.

In de AS-IS situatie is dit een inrichting zonder gebruik te maken van CC en voor de TO-BE situatie is dit ingericht gebruik makend van CC, afhankelijk van de keuze voor het Cloud Service en Cloud Deployment Model vanuit de organisatiedoelstellingen.

Door de AS-IS en TO-BE situatie tegen elkaar af te zetten kunnen we vervolgens impact bepalen.

#### 4.6.2 Vertaling naar aandachtspunten en toetsing aan interviewresultaten

Uitgaande van het in sectie 4.6.1 geconstrueerd schema zijn in onderstaande tabel in de 2<sup>e</sup> kolom de relevante aandachtspunten vanuit de theorie afgeleid. Deze aandachtspunten zijn vervolgens gerelateerd aan de uitkomsten uit de interviews m.b.t. de deelgebieden. Daarbij is getracht in de 3<sup>e</sup> kolom aan te geven in hoeverre een aandachtspunt vanuit de praktijk wordt onderschreven, in de 4<sup>e</sup> kolom aangevuld met een onderbouwing vanuit de interviews.



Nr	Aandachtspunt, afgeleid uit de theorie	Onderkend vanuit de praktijk	Motivatie vanuit de interviews
1	Doelstellingen van de organisatie zijn inzichtelijk	Ja	Interviews bevestigen dat DQM, BDA en inzet van CC begint bij doelstellingen van de organisatie. DQM, BDA en CC moeten niet beschouwd worden als doelen op zich
2	Doelstellingen van de organisatie zijn vertaald naar doelstellingen voor DQM	Ja	Interviews bevestigen dat doelstellingen voor DQM zijn afgeleid uit de doelstellingen van de organisatie rekening houdend met het vereiste kwaliteitsniveau voor data.
3	Doelstellingen voor DQM zijn vertaald naar eisen voor datakwaliteit uitgedrukt in de dimensies kwaliteitscriteria, eigenschappen van BD en bijbehorende toleranties	Ja	Uit de interviews komt naar voren dat doelstellingen voor DQM zich vertalen naar kwaliteitseisen voor data. Daarbij wordt opgemerkt dat het beter is om te spreken over kwaliteitsaspecten/criteria die, afhankelijk van het vraagstuk wat speelt, doorvertaald worden naar eisen aangaande deze aspecten/criteria. Uit de interviews komt tevens naar voren dat in de kwaliteitsrekening moet worden gehouden met de eigenschappen van BD en toleranties. Toleranties zijn daarbij afhankelijk van het vraagstuk wat speelt en impact van bijbehorende beslissing
4	Doelstellingen voor DQM en daarvan afgeleide eisen voor datakwaliteit zijn vertaald naar beleid en standaarden voor datakwaliteit	Ja	Het belang van doorvertaling naar beleid en standaarden wordt vanuit de interviews bevestigd.
5	Stappen van de AS-IS inrichting van de Big Data Pipeline op niveau van de business laag zijn inzichtelijk	Twijfel	Interviews geven hier een wisselend beeld. Een aantal respondenten geeft aan dat BDA zich met name zal toespitsen op de Information Systems laag en niet op de Business Laag
6	Beleid en standaarden voor datakwaliteit zijn doorvertaald naar een AS-IS proces- en organisatieinrichting voor de Big Data Pipeline met aandacht voor maatregelen voor toetsing en verbetering van datakwaliteit	Twijfel	Interviews geven hier een wisselend beeld. Belang van een AS-IS situatie wordt onderkend maar de vraag is of dit zich voor BDA en bijbehorende maatregelen voor DQM manifesteert op de Business laag of meer tot uitdrukking komt in de Informatie Systems laag. Aantal respondenten bevestigen maatregelen op niveau van de Business Laag terwijl anderen dit meer toeschrijven naar de Information Systems laag. Daarnaast is aangegeven dat een EA mogelijk een te hoog/abstract niveau is.
7	Doelstellingen van de organisatie zijn doorvertaald naar een keuze voor het te hanteren Cloud Service Model en Cloud Deployment Model	Ja	Uit de interviews is naar voren gekomen dat dit de weg is die bewandeld zou moeten worden. CC moet niet gezien worden als een doel op zich.
8	Stappen van de TO-BE inrichting van de Big Data Pipeline op niveau van de business laag zijn inzichtelijk rekening houdend met het te Cloud Service Model en Cloud Deployment Model	Twijfel	Interviews geven hier een wisselend beeld. Een aantal respondenten geeft aan dat BDA zich met name zal toespitsen op de Information Systems laag en niet op de Business Laag
9	In de TO-BE proces en organisatieinrichting van de Big Data Pipeline is bij de maatregelen voor toetsing en verbetering van datakwaliteit rekening gehouden met het gekozen Cloud Service Model en Cloud Deployment Model	Twijfel	In aansluiting op punt 6 geven de interviews een wisselend beeld ten aanzien van de positionering van BDA en bijbehorende DQM maatregelen op het niveau van de business laag en het nut van een EA. In de interviews wordt wel het beeld bevestigd dat de keuze van het Cloud Deployment Model en Cloud Service Model zijn effect heeft op de onderdelen van de diverse lagen van een EA.
10	De TO-BE proces- en organisatieinrichting is in lijn met beleid en standaarden voor datakwaliteit	Ja	Interviews geven aan dat beleid en standaarden voor datakwaliteit zijn doorwerking hebben in de proces- en organisatieinrichting

Figuur 19: Onderkende aandachtspunten vanuit de theorie incl. toetsing aan de praktijk

Uit bovenstaande tabel is op te maken dat niet alle aandachtspunten vanuit de theorie volledig onderschreven worden vanuit de praktijk. Aandachtspunten gericht op het inrichtingsonafhankelijk deel voor DQM worden onderkend. Er ontstaat echter een wisselend beeld, kijkend naar de toepassing van een EA voor de weergave van de inrichting van BDA en bijbehorende maatregelen voor DQM.

Bovenstaande moet wel in het licht beschouwd worden van de beginfase waarin de organisatie zich bevindt wat betreft de invoering van BDA en de ontwikkelpunten die de organisatie zelf ziet m.b.t. DQM.

Naast bovengenoemde aandachtspunten vanuit de theorie zijn uit de interviews ook een aantal aanvullende aandachtspunten te herleiden die niet direct te koppelen zijn aan de inzichten verkregen vanuit de theorie. Deze zijn weergegeven in onderstaande tabel.

Nr	Aanvullend Aandachtspunt	Motivatie vanuit de interviews
11	Bewaak hergebruik van maatregelen voor toetsing van datakwaliteit	Respondenten geven aan dat door omzetting van Big Data naar een andere vorm waarvoor reeds technieken voor toetsing van datakwaliteit zijn ingericht, hergebruik van deze technieken kan worden bewerkstelligd
12	Maak onderscheid in gegevensbeheer en gegevensgebruik	Respondenten geven aan dat bij de proces- en organisatieinrichting ter borging van de datakwaliteit rekening moet worden gehouden met het onderscheid tussen gegevensgebruik en gegevensbeheer. Leg de verantwoordelijkheid voor het voorzien in het gewenste kwaliteitsniveau bij de bronhouder.
13	Hou rekening met eigenaarschap van data en bijbehorende ketensamenwerking	Respondenten geven aan dat organisaties gebruik kunnen maken van externe gegevensbronnen waarbij het eigenaarschap en verantwoordelijkheid voor datakwaliteit bij een externe partner is belegd
14	Zorg dat het gebruik van EAM en EA organisatie-breed geadopteerd is	Alhoewel bij respondenten het nut van een EAM en een EA onderkend wordt, is het gebruik en nut hiervan nog niet organisatie-breed gecommuniceerd en gedeeld. Daarbij wordt opgemerkt dat het gebruik van een EAM en EA voor DQM en BDA nog niet door alle respondenten gedeeld wordt.
15	Zorg dat in het gebruik van EAM en EA aandacht wordt besteed aan de zachte (veranderkundige) kant	Respondenten geven aan dat het risico bestaat dat het gebruik van EAM en een EA te instrumenteel benaderd wordt. In migratietrajecten is het van belang aandacht te besteden aan veranderkundige aspecten in lijn met het verandervermogen van de organisatie en betrokken medewerkers

*Figuur 20: Aanvullende aandachtspunten vanuit de praktijk*

## 5. Discussie, conclusies en aanbevelingen

### 5.1 Discussie – reflectie

Het uitgevoerde onderzoek is gebaseerd op een theoretische verkenning van 4 deelgebieden, overeenkomend met de 4 deelvragen, resulterend in termen en onderlinge relaties waarvan de herkenbaarheid is getoetst aan de hand van een praktijkcasus. Ter beantwoording van de hoofdvraag zijn de resultaten uit de theoretische verkenning bij elkaar gebracht en vertaald naar een lijst met aandachtspunten die vervolgens zijn geprojecteerd op de bevindingen uit de praktijkcasus. Ten aanzien van de verkregen resultaten en het daarbij gevolgd proces zijn echter enkele kritische kanttekeningen te plaatsen.

#### *Gevolgd proces*

Kijkend naar het gevolgde proces is in het onderzoek getracht, o.b.v. verkenning van de deelgebieden, bovenliggende hoofdvraag te beantwoorden. Gebleken is dat de afzonderlijke deelgebieden op zichzelf reeds een onderzoeksgebied kunnen vormen, waardoor het risico werd gelopen dat het onderzoek te groot werd en haalbaarheid in het geding kwam. Door de deelgebieden te benaderen in de context van de hoofdvraag met daarmee gepaard gaande selectie van onderwerpen, is getracht dit risico in te perken.

Bij de theoretische verkenning per deelgebied zijn de beschikbare artikelen vanuit het literatuuronderzoek in detail bekeken en zijn, o.b.v. een persoonlijke interpretatie, relevante onderdelen eruit gehaald en vertaald naar termen en hun onderlinge relaties. Door deze persoonlijke interpretatie wordt de relatie met onderliggende artikelen enigszins doorbroken (zie verder onder reflectie op betrouwbaarheid).

De resultaten uit de theoretische verkenning hebben de grondslag gevormd voor de uitgevoerde interviews. Alhoewel deze grondslag afbakenend en richtinggevend heeft gewerkt, wordt hier echter ook een risico gelopen dat teveel naar bepaalde resultaten toegewerkt wordt (zie verder onder reflectie op betrouwbaarheid). Geleverde antwoorden tijdens de vragen waren voor een groot deel afhankelijk van de mate van volwassenheid van de organisatie en bekendheid/expertise van de geïnterviewden in het betreffende domein (zie verder onder reflectie op externe validiteit).

De thematische analyse o.b.v. de interviewverslagen is gedaan aan de hand van de termen zoals deze vanuit de theoretische verkenning zijn onderkend. Naast de afzonderlijke termen is in de tekstfragmenten ook gekeken naar herkenning van relaties tussen termen. Dit laatste onderdeel is echter maar ten dele gelukt omdat de resultaten uit de interviews hier maar beperkt uitsluitsel in gaven.

Beantwoording van de hoofdvraag vanuit een theoretisch oogpunt heeft plaats gevonden door relevante termen en onderlinge relaties uit de deelgebieden te selecteren/te combineren en deze vervolgens te vertalen naar een lijst met aandachtspunten. Hieraan ligt een persoonlijke interpretatie aan te grondslag die de reproduceerbaarheid onder druk zet (zie verder onder reflectie op betrouwbaarheid). Door de aandachtspunten vervolgens te toetsen aan de resultaten uit interviews is getracht de volledigheid/juistheid van de aandachtspunten te beoordelen vanuit een praktisch oogpunt.

#### *Bereikte resultaten*

Kijkend naar de bereikte resultaten kan het volgende opgemerkt worden.

De aandachtspunten zoals deze vanuit een theoretisch oogpunt zijn onderkend, worden voor meer dan de helft onderschreven vanuit een praktisch oogpunt. Bovendien zijn een aantal extra aandachtspunten vanuit de interviews naar voren gekomen. Daarbij moet wel worden opgemerkt

dat deze resultaten tot stand zijn gekomen o.b.v. een toets bij één casusorganisatie. Een aantal punten zijn in dat kader vermeldenswaardig.

De toegevoegde waarde van een EA voor BDA en DQM wordt niet helemaal gedeeld. Argumentatie hiervoor is dat een EA te hoog over is en BDA en DQM voor een groot deel landen in de Informations Systems laag. Een tegenargument hiervoor zou kunnen zijn dat onderdelen van een EA verder uitgewerkt kunnen worden in deelarchitecturen waarvoor een EA de kapstok vormt. Ook vanuit de literatuur wordt dit onderkend waarbij een EA in samenhang met meer low-level modelleringsstandaarden moet worden beschouwd (Gill, A.Q., 2015). Daarnaast kan bij de inrichting van zowel DQM als BDA onderscheid worden gemaakt in het proces en bijbehorende verantwoordelijkheden (als onderdeel van de Business laag) en de wijze waarop dit proces middels toepassingen ondersteund wordt (als onderdeel van de Information Systems laag).

Een ander punt vormt de mate waarin een EAM en EA in zijn algemeenheid gebruikt worden. Een aantal respondenten geven aan dat het gebruik van een EAM en EA nog geen gemeengoed is binnen de organisatie en communicatie over EAM/EA en de toegevoegde waarde hiervan nog beter over het voetlicht moeten worden gebracht. Adoptie en gebruik van EAM/EA in zijn algemeenheid is een herkenbaar thema waaraan reeds diverse onderzoeken aandacht aan hebben besteed (Dang, D.D., Pekkola, S., 2017) (Niemi, E., Pekkola, S., 2019).

Daarnaast wordt opgemerkt dat in het gebruik van EAM ook aandacht moet worden besteed aan het zachte (veranderkundige) aspect. Het beheersen van veranderingen betekent namelijk omgaan met weerstand. Veranderingen brengen namelijk onzekerheid en discomfort met zich mee die weerstand tot gevolg hebben. Een EAM dient hier rekening mee te houden (Espinoza, F., 2007).

Tenslotte vormt de vertaling van doelstellingen van een organisatie naar een Cloud Service en Deployment Model en bijbehorende invoering een punt van aandacht. Alhoewel het gedachtegoed gedeeld wordt, vormt het verkrijgen van draagvlak, gekoppeld aan een implementatie een niet te onderschatten inspanning waarin EAM een rol kan spelen mits aandacht wordt besteed aan het reeds genoemde veranderkundige aspect. Dit aspect is ook vanuit de literatuur herkenbaar waarbij change management en organisatiecultuur als belangrijke factoren onderkend worden bij het gebruik van CC (Gangwar, H., 2017).

#### *Reflectie op interne validiteit*

Voor elk van de 4 onderkende deelgebieden zijn meerdere medewerkers geïnterviewd waarbij de medewerkers verschillende rollen binnen de organisatie hebben. Middels deze vorm van triangulatie is getracht de interne validatie te waarborgen. Tijdens de interviews is echter gebleken dat, kijkend over de verschillende interviews heen, voor gelijksoortige vragen niet altijd gelijksoortige antwoorden zijn geleverd, hetgeen de interne validiteit toch enigszins onder druk zet. Daarbij kan de mogelijke kennis/ervaring van de geïnterviewde met het deelgebied een onderliggende oorzaak vormen.

Voor twee deelgebieden zijn tevens achtergronddocumenten aangereikt waarvan één document niet in overeenstemming is met de actualiteit en een ander document niet tot invoering heeft geleid. Daarmee is de waarde van deze achtergronddocumenten bij onderbouwing van de geleverde antwoorden in de interviews beperkt.

#### *Reflectie op externe validiteit:*

De casusorganisatie betreft een organisatie die een brede, meerjarige, achtergrond heeft in het uitvoeren van data-analyses maar zich wat betreft het toepassen van BDA nog in een beginfase bevindt en voor DQM nog ontwikkelpunten ziet. Bij de casusorganisatie is niet expliciet sprake van een strategie voor CC die vervolgens vertaald is naar een invoeringstraject. Daarnaast is bij de geïnterviewden slechts beperkte praktische expertise aanwezig inzake BDA en is binnen de organisatie geen sprake van zuivere migratie van BDA vanuit een On Premise situatie naar een CC ondersteunde situatie. Dit geheel heeft het onderzoeksresultaat in zekere mate beïnvloed en zet druk op de generaliseerbaarheid van de bereikte resultaten.

### *Reflectie op betrouwbaarheid*

Op basis van literatuur is getracht per deelgebied te komen tot aandachtspunten uitgedrukt in termen en hun onderlinge relatie. Betrokken artikelen zijn resultaten uit een uitgevoerde literatuurstudie waarvan de werkwijze met gehanteerde zoektermen, zoekmachine, filter- en sorteercriteria is verwoord in hoofdstuk 2. Hiermee is getracht een bepaalde reproduceerbaarheid te garanderen.

Aan de vertaling van de inhoud van de artikelen naar de relevante termen per deelgebied en hun onderlinge relaties ligt echter een persoonlijke interpretatie ten grondslag hetgeen de reproduceerbaarheid onder druk zet.

Ook ten aanzien van de samenstelling van de aandachtspunten vanuit theorie ter beantwoording van de hoofdvraag ligt, alhoewel de gehanteerde werkwijze is verwoord, een persoonlijke interpretatie aan ten grondslag.

Bij de aanpak van de praktijkcasus is gekozen voor semi-gestructureerde interviews waarbij een vragenlijst een leidraad vormt. Deze vragenlijst is opgezet per deelgebied en gebaseerd op de inzichten die vanuit het theoretisch deel zijn verkregen. Gegeven de kanttekeningen die zijn geplaatst ten aanzien van het theoretisch deel heeft dit ook zijn effect op de reproduceerbaarheid van de vragenlijst.

Door de vragen als leidraad te gebruiken voor het interview is sprake van een bepaalde mate van sturing in het leveren van antwoorden waardoor de betrouwbaarheid van het onderzoek enigszins onder druk komt te staan.

Resultaten uit de interviews zijn teruggekoppeld in de vorm van een interviewverslag met het verzoek te reageren. Alle 8 respondenten hebben gereageerd. Het overgrote deel was het eens met hetgeen in het verslag verwoord is. 4 respondenten hadden kleine tekstuele aanvullingen of correcties. Daarmee vormen de verslagen een betrouwbare weergave van hetgeen in de interviews is besproken.

Gebruikte codering voor thematische analyse van de tekstfragmenten uit de interviewverslagen is gebaseerd op termen die vanuit de theorie zijn verkregen en aangevuld met termen die vanuit de interviews zijn geconstateerd. Wijze waarop tekstfragmenten zijn gekoppeld aan codes is gebaseerd op een persoonlijke interpretatie waardoor de reproduceerbaarheid ook hier onder druk staat.

### *Reflectie op ethische aspecten*

Ten aanzien van de ethische aspecten wordt opgemerkt dat ten aanzien van de, in sectie 3.4 benoemde, punten, voor zover inzichtelijk, geen belemmeringen zijn geconstateerd.

Interviewverslagen zijn alleen teruggekoppeld naar de geïnterviewden en voor verdere verwerking zijn de resultaten geanonimiseerd.

## 5.2 Conclusies

De onderzoeksresultaten zijn vertaald naar resultaten per deelgebied die vervolgens bij elkaar zijn gebracht ter beantwoording van de hoofdvraag.

De onderzoeksresultaten van de deelgebieden leveren de volgende conclusies.

Gezien het stijgend belang van data en bijbehorende analyses in het realiseren aan doelstellingen van organisaties wordt het belang van DQM onderstreept. Het daarbij vereist kwaliteitsniveau is afhankelijk van de (maatschappelijke) vraagstukken waarop de organisatie een antwoord moet geven. Doorvertaling naar onderdelen van een EA vereist echter als randvoorwaarde dat het gebruik van een EA in zijn algemeenheid en toepasbaarheid voor het DQM domein, organisatie-breed, is geaccepteerd.

Ook binnen het BDA domein wordt het belang van DQM onderkend. Kwaliteit van de analyse resultaten zijn immers afhankelijk van het resultaat van de grondstoffen (lees de data) die hiervoor



wordt ingezet. De vertaling naar de inrichting van BDA en bijbehorende maatregelen voor borging van datakwaliteit, uitgedrukt in een EA, hangt hier echter ook weer samen met de mate waarin het gebruik van een EA en toepasbaarheid hiervan voor BDA en DQM binnen de organisatie gedragen wordt.

Ten aanzien van de ondersteuning middels CC wordt het gebruik van een Cloud Service Model en Cloud Deployment Model, afgeleid uit de organisatiedoelstellingen, onderschreven. Wijze waarop dit een vervolg krijgt in de uitwerking op een EA en relevante componenten voor BDA en DQM hangt ook weer samen met hetgeen eerder is opgemerkt ten aanzien van het gebruik van een EA.

De weergave van een AS-IS inrichting (zonder CC) en een TO-BE inrichting (met CC), gevolgd door het vergelijk van deze twee inrichtingen om tot impactbepaling te komen maakt EAM een toepasbaar instrument. EAM reikt echter verder dan het weergeven van een AS-IS en TO-BE situatie. Het betreft ook de doorvertaling naar een Roadmap, uitvoering en monitoring van projecten (zowel inhoudelijk als procesmatig). Toepassen van EAM begint echter met, de organisatie-brede adoptie, van EAM. Daarnaast vormt aandacht voor het veranderkundig aspect, rekening houdend met het verandervermogen van de organisatie, een belangrijke randvoorwaarde voor het succesvol inzetten van EAM.

Geprojecteerd op de hoofdvraag kan het volgende geconcludeerd worden.

Zinnvolle toepassing van EAM voor impactbepaling bij migratie van BDA naar de Cloud begint bij de acceptatie van EAM in zijn algemeenheid. Onderdeel van een EAM is een EA. Los van adoptie van het gebruik van een EA in zijn algemeenheid speelt hier de adoptie van de wijze waarop BDA en DQM worden uitgedrukt in een EA. Voor de AS-IS situatie is dit een inrichting zonder CC en voor de TO-BE met CC, waarbij in de TO-BE situatie rekening is gehouden met de keuze in CC ondersteuning, uitgedrukt in een Cloud Service en Deployment Model, afgeleid van organisatiedoelstellingen. Bovenstaande kan uitgedrukt worden in een aantal aandachtspunten, weergegeven in sectie 4.6.2, waarmee de organisatie rekening moet houden, wil EAM zinvol worden ingezet.

### 5.3 Aanbevelingen voor de praktijk

Organisaties die overwegen BDA onder te brengen in de Cloud kunnen bij impactbepaling voor DQM EAM als methodiek inzetten. De lijst met aandachtspunten, zoals weergegeven in sectie 4.6.2, kan hierbij als handvat gehanteerd worden.

Het migratievraagstuk raakt echter een aantal deelgebieden waarbij kritisch moet worden gekeken naar mate van volwassenheid en opgebouwde expertise binnen deze deelgebieden. Risico bestaat dat het ambitieniveau te hoog is, gegeven de stand van zaken op de deelgebieden, waardoor de haalbaarheid van het migratietraject als geheel in gevaar komt. Geadviseerd wordt dan ook eerst deze deelgebieden onder loep te nemen en daarin reeds verbeteringen te zoeken. Hierbij kan gedacht worden aan de volgende punten:

- Mate waarin DQM in zijn algemeenheid is ingevoerd en de wijze waarop dit vervolgens is toegepast binnen het BDA domein
- Mate van adoptie en gebruik van een EA en EAM binnen de organisatie in zijn algemeenheid t.b.v. het doorvoeren van veranderingstrajecten
- Mate van adoptie van het gebruik van een EA t.b.v. de inrichting van DQM en BDA.
- Mate waarin voorzien is in een Cloudstrategie vertaald naar een Cloud Service en Cloud Deployment Model en de wijze waarop EA en EAM wordt ingezet voor het implementatie van deze strategie.

## 5.4 Aanbevelingen voor verder onderzoek

Het onderzoek is uitgesplitst naar deelgebieden waarvan de resultaten vervolgens bij elkaar zijn gebracht ter beantwoording van de hoofdvraag.

Ten eerste wordt opgemerkt dat op de afzonderlijke deelgebieden en relaties tussen deelgebieden nog nader onderzoek gerechtvaardigd is. Enkele onderzoeksonderwerpen die hierbij genoemd kunnen worden, zijn:

- Wijze waarop de kwaliteit van data analyse resultaten uit een BDA traject geborgd kan worden en zich vertaalt naar de inrichting van een Big Data Pipeline.
- Nut van het gebruik van een EA ter borging van datakwaliteit
- Wijze waarop de verschillende mogelijke invullingen van een Cloud Service Model en Cloud Deployment Model doorwerken in de lagen van een EA
- Scenario's voor implementatie van een Cloudstrategie binnen organisaties en de wijze waarop een EAM hiervoor ingezet kan worden

Ten tweede wordt opgemerkt dat de onderzoeksresultaten zijn gebaseerd op de kennis/ervaring binnen één casusorganisatie. Om een goed gefundeerd antwoord te geven op de hoofdvraag is het noodzakelijk het onderzoek uit te breiden naar meer organisaties, opererend in verschillende sectoren. Daarbij moet kritisch gekeken naar de mate van volwassenheid en opgebouwd expertiseniveau op het gebied van BDA, DQM, CC en het gebruik van EA/EAM.

## Bibliografie

- Ahleman F., Stettiner E., Messerschmidt, M., Legner, C. (2012). *Strategic Enterprise Architecture Management*. Heidelberg: Springer.
- Al-Ruithe, M., Benkhelifa, E., Hameed, K. (2018). Data Governance Taxonomy: Cloud versus Non-Cloud. *Sustainability*.
- Association of Universities in the Netherlands (VSNU). (2014). *The Netherlands Code of Conduct for Academic Practice*. The Hague.
- Bernal, W.N., Cabalero, G.C., Sanchez, J.O., Paez-Logreira, H. (2016). Enterprise Architecture Framework Oriented to Cloud Computing Services. *International Conference on Computers Communications and Control*.
- Burmeister, F., Drews, P., Schirmer, I. (2018). Towards an extended enterprise architecture meta-model for big data - a literature-based approach. *Twenty-fourth America Conference on Information Systems*, (pp. 1-10). New Orleans.
- Cecilia, A., Rusli, A., Rodziah, A., Yah, J. . (2017). Conceptual model development of big data analytics implementation assessment effect on decision making. *International Journal of interactive media and Artificial Intelligence*, 101-106.
- Chen, H., Chiang, R.H.L., Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Quarterly*.
- Comuzzi, M., Patel, A. (2016). How organisations leverage Big Data: a maturity model. *Industrial Management & Data Systems*, 1468-1492.
- Dang, D.D., Pekkola, S. . (2017). Systematic Literature Review on Enterprise Architecture in the Public Sector. *The Electronic Journal of e-Government*.
- Demchenko, Y., Laat de, C. (2014). Defining Architecture Components of the Big data Ecosystem. *Proceedings of the 2014 International Conference on Collaboration Technologies and Systems*. Piscataway, NJ: IEEE.
- Di Martino, B., Aversa, R., Cretella, G., Esposito, A. (2014). Big Data (lost) in the cloud. *Int. J. Big Data Intelligence*.
- Dong, X., Heng, H., Chao, L., Yongchuan, L., Houbo, X. (2018). *Scene-Bases Big Data Quality Management Framework*. Singapore: Springer.
- Espinoza, F. (2007). Enterprise Architecture and Change Management . *Journal of Enterprise Architecture*.
- Gampfer F., Jürgens A., Müller M., Buchkremer, R. (2018). Past, current and future trends in enterprise architecture - A view beyond the horizon. *Computers in Industry*.
- Gangwar, H. (2017). Cloud computing usage and its effect on organizational performance. *Human Systems Management*.
- Gantz J., Reinsel, D. (2012). The digital universe in 2020: big data, bigger digital shadows and biggest growth in the far east. *IDC iView IDC Analyse the Future*.

- Gill, A.Q. (2015). Agile enterprise architecture modelling: Evaluating the applicability and integration of six modelling standards. *Information and Software Technology*.
- Gong, Y., Janssen, M. (2020). Roles and Capabilities of Enterprise Architecture in Big Data Analytics Technology Adoption and Implementation. *Journal of Theoretical and Applied Commerce Research*, 37-51.
- Harrison, T.M, et al. (2019). The Data Firehose and AI in Government: Why Data Management is a Key to Value and Ethics. *20th annual international conference on digital government research* (pp. 171-176). Dubai: ACM.
- Hashem, I.A.T., et al. (2014). The rise of "big data" on cloud computing: review and open research issues. *Information Systems*.
- Khalilijafarabad, A., Helfert, M., Mouzhi, G. (2016). Developing a Data Quality Research Taxonomy - an Organizational Perspective. *ICIQ*.
- Kim, H.Y., Cho, J-S. (2017). Data Governance framework for big data implementation with NPS Case Analysis in Korea. *A Journal of the Academy of Business and Retail Management*, 36-46.
- Kühne, B., Bornholt, J., Zolnowski, A., Böhm, T. (2019). Making Data Tangible for Data-driven Innovations in a Business Model Context. *Twenty fifth Americas Conference on Information Systems*. Cancun.
- Lnenicka, M., Komarkova, J. (2018). Developing a government enterprise architecture framework to support the requirements of big and open linked data with the use of cloud computing. *International Journal of Information Management*.
- Lnenicka, M., Machova, R., Komarkova, J., Cermakova, I. (2017). Components of Big Data Analytics for Strategic Management of Enterprise Architecture.
- Mahmood, Z., Hill, R. (2011). *Cloud Computing for Enterprise Architectures*. London: Springer.
- Mell, P., Grance, T. (2011). The NIST definition of cloud computing. Recommendations from the National Institute of Standards and Technology. *NIST special publ.*
- Merino, J., Caballero, I., Rivas, B., Serrano, M., Piattini, M. (2015). A Data Quality in Use model for Big Data. *Future Generation Computer Systems*.
- Mikalef, P., Pappas, I.O., Krogstie, J., Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Inf. Syst E-Bus Manage*.
- Moreno C., C. R.-V. (2019). Data and Artificial Intelligence Strategy: A Conceptual Enterprise Big Cloud Architecture to Enable Market-Oriented Organisations. *Special issue on Use Cases of Artificial Intelligence, Digital Marketing and Neuroscience*.
- Niemi, E., Pekkola, S. (2019). The Benefits of Enterprise Architecture in Organizational Transformation. *Bus Inf Syst Eng*.
- Ohbyung, K., Namyoon, L., Bongsik, S. (2014). Data Quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*.
- Otto, B. (2017). *Conceptualizing the knowledge to manage and utilize data assets in the context of digitization: Case studies of multinational Industrial Enterprises*. Dortmund: Fraunhofer Innovation Center for Logistics and IT.

- Otto, B., Yang, Y.L., Caballero, I. (2011). Information and data quality in business networking: a key concept for enterprises in its early stages of development. *Electron Markets*.
- Panian, Z. (2016). Some practical experiences in Data Governance. *World Academy of Science, Engineering and Technology*.
- Peppers, K., Tuunanen, T., Rothenberger, M.A, Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*.
- Popovic, A., Hackney, R., Tassabehji, R., Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. *Inf System Front*.
- Runeson, P., Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empir Software Eng*.
- Saunders, M.K.N, Lewis, P., Thornhill, A. (2019). *Research methods for business students - eighth edition*. Harlow: Pearson.
- Schmidt, R., Möhring, M. (2013). Strategic alignment of Cloud Based Architectures for Big Data. *IEEE International Enterprise Distributed Object Computing Conference Workshops*.
- Sharma, S. (2016). Expanded cloud plumes hiding Big Data ecosystem. *Future Generation Computer Systems*.
- Taleb, I., Serhani, M.A., Rchida, D. (2018). Big Data Quality Assessment Model for Unstructured Data.
- Weber, K., Otto, B., Osterle, H. (2009). One size does not fit all - a contingency approach to data governance. *ACM Journal of Data and Information Quality*.

## Bijlage I: Lijst met afkortingen

BD:	Big Data
BDA:	Big Data Analytics
CC:	Cloud Computing
DQ:	Data Quality
DQM:	Data Quality Management
EA:	Enterprise Architecture
EAM:	Enterprise Architecture Management

## Bijlage II: Resultaten Samenhang DQM met BDA

Deze bijlage geeft voor de samenhang tussen DQM en BDA inzicht in:

- Onderkende termen vanuit de theoretische verkenning
- Geformuleerde vragen die als leidraad golden voor de gevoerde interviews
- Resultaten uit de gevoerde interviews en daaraan gekoppelde analyse
- Samenvatting van de analyse resultaten

### Onderkende termen

O.b.v. de verkregen inzichten vanuit een theoretisch oogpunt zoals weergegeven in sectie 4.2.1 zijn de volgende termen te onderkennen.

Termen
Doelstelling Organisatie
Doelstelling DQM
Eisen Datakwaliteit
Eigenschappen BD
Kwaliteitscriteria
Toleranties
Toetsen Datakwaliteit
Verbeteren Datakwaliteit

Figuur 21: Termen samenhang DQM met BDA

### Geformuleerde vragen

Op basis van het verkregen theoretisch inzicht en daaraan gekoppelde termen zijn onderstaande vragen geformuleerd.

Nr. Vraag	Vraag
A1	Wat zijn de doelstellingen die de organisatie heeft ten aanzien van DQM
A2	Wat zijn de doelstellingen die de organisatie heeft ten aanzien van BDA
A3	Op welke wijze zijn deze gekoppeld aan de overkoepelende organisatiedoelstellingen
A4	Op welke wijze zijn de doelstellingen voor DQM vertaald naar eisen voor data kwaliteit
A5	Welke eisen voor data kwaliteit worden onderkend
A6	Welke kenmerken voor Big Data worden onderkend (wat karakteriseert Big Data)?
A7	Hoe zijn de eisen voor data kwaliteit gekoppeld aan deze kenmerken
A8	Uitgaande van de kenmerken van Big Data, welke toleranties (goed is goed genoeg) voor data kwaliteit worden hierbij geaccepteerd in lijn met de doelstellingen die met BDA nagestreefd worden
A9	Op welke wijze zijn doelstellingen en kwaliteitseisen binnen de organisatie vertaald naar: <ul style="list-style-type: none"><li>• Toetsing (Assessment) van de data kwaliteit</li><li>• Verbetering (corrigerende acties) van de data kwaliteit</li></ul>

Figuur 22: Vragen samenhang DQM met BDA

### Resultaten uitgevoerde interviews en daaraan gekoppelde analyses

In onderstaande tabel zijn de antwoorden op de gestelde vragen weergegeven waarbij o.b.v. de tekst is gekeken aan welke term of relatie van termen dit te relateren is. Een antwoord kan hierbij opgeknipt zijn in meerdere tekstfragmenten om het vervolgens te kunnen relateren aan de termen en onderlinge relaties. Een groot deel van de termen en onderlinge relaties hebben hun oorsprong in het verkregen theoretisch inzicht. Een deel van de termen en onderlinge relaties kunnen aanvullend geïntroduceerd zijn n.a.v. de gegeven antwoorden.



Toelichting kolommen uit de tabel:

- Nr. Vraag: Link naar de tabel met geformuleerde vragen (het nummer correspondeert met het nummer in de tabel met vragen)
- Nr. Antwoord: Identificatie van een tekstfragment van het antwoord op een vraag
- Interview: Identificatie van het interview
- Tekst: Tekstfragment van een antwoord wat te koppelen is aan een term of relatie tussen termen
- Term of relatie tussen termen: Termen of relatie tussen termen die herkend kunnen worden vanuit het tekstfragment (bij relatie tussen termen zijn de termen gescheiden door een “-”)

Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
A1	1	I-1	Voorzien in zodanige datakwaliteit dat op basis hiervan de juiste beslissingen genomen kunnen worden rekening houdend met de genoemde piramide (van data naar informatie naar kennis naar actie).	Doelstelling Organisatie - Doelstelling DQM
A1	2	I-1	Indien de data van onvoldoende kwaliteit is, leidt dit uiteindelijk tot minder goede beslissingen	Doelstelling Organisatie - Kwaliteitseisen
A1	3	I-1	Dat wil niet zeggen dat je altijd een volledige dataset als basis moet nemen. Je kunt werken met een steekproef mits deze steekproef wel een goede weerspiegeling is van de volledige set.	Toetsen Datakwaliteit
A1	4	I-2	Qua inhoud voldoen aan de kwaliteitseisen die vanuit Beleid gesteld worden waarbij Gegevensmanagement adviserend is	Doelstelling DQM
A1	5	I-2	Ervoor zorgen dat technisch gezien de data op orde is: in lijn met het datamodel en metadata en voldoen aan integriteitsregels	Doelstelling DQM
A1	6	I-4	Zelf niet bekend met deze doelstellingen op het van DQM en BDA. Dat wil overigens niet zeggen dat ze er niet zijn	Doelstelling DQM; Doelstelling BDA
A1	7	I-4	Gegevensmanagement zit tussen de leverancier en de gebruiker. Gebruiker heeft last als de data inhoudelijk niet klopt. Gegevensmanagement heeft er last van als de data technisch niet klopt	Doelstelling Organisatie - Kwaliteitseisen
A1	8	I-5	We hanteren een aantal principes: zoals data bij de bron houden zodat we dit consistent te houden	Doelstelling Organisatie - Doelstelling DQM
A1	9	I-5	We maken gebruik van een GLP (GegevensLeveringsProtocol) zodat we gegevens op een eenduidige en juiste manier geleverd krijgen met controles tegen metadata en of geleverd wordt conform het GLP In GLP wordt het formaat afgesproken en de eigenschappen waaraan de levering moet voldoen	Kwaliteitseisen
A1	10	I-5	Door het GLP te op te stellen kun je het ook goed toetsen. Cruciaal voor het verdere gegevensgebruik	Toetsen Datakwaliteit
A2	11	I-2	BDA is een hype in zit voornamelijk nog in de innovatie sferen.	Doelstelling BDA
A2	12	I-2	BDA Bezien vanuit Top-Down wordt het gezien als een doel op zich	Doelstelling BDA
A2	13	I-2	BDA Bezien van Bottom-Up kan het helpen bestaande vraagstukken effectiever en efficiënter op te lossen	Doelstelling Organisatie - Doelstelling BDA
A2	14	I-2	Op een effectievere en efficiëntere manier beantwoorden aan maatschappelijke vraagstukken	Doelstelling Organisatie
A2	15	I-2	Bijdrage leveren in de beleidsvorming door dicht bij de burger te zitten qua actualiteit van gegevens en reacties hierop	Doelstelling Organisatie - Kwaliteitseisen
A2	16	I-2	BDA heeft de potentie een grote bijdrage te leveren in het oplossen van maatschappelijke vraagstukken	Doelstelling Organisatie - Doelstelling BDA
A2	17	I-2	BDA is echter geen doel op zich en het is niet zo dat met 1 druk op de knop de verandering is doorgevoerd.	Doelstelling BDA
A2	18	I-2	BDA heeft impact op de organisatie en betekent een trendbreuk met verleden in de beantwoording van beleidsvraagstukken en heeft effect op de wijze waarop je monitort	Doelstelling Organisatie - Doelstelling BDA
A3	19	I-1	Organisatie willen hun prestaties verbeteren.	Doelstelling Organisatie
A3	20	I-1	Om hiertoe in staat te zijn moeten ze de juiste beslissingen nemen.	Doelstelling Organisatie
A3	21	I-1	Om tot de juiste beslissingen te komen moet de organisatie zich baseren op kwalitatief goede data.	Doelstelling Organisatie - Kwaliteitseisen
A3	22	I-2	Formulering van kwaliteitseisen gaat bottom up. Er is geen bovenliggende kapstok	Doelstelling Organisatie - Kwaliteitseisen
A4	23	I-2	Daar bovenop moet beleid richting geven aan aanvullende kwaliteitseisen	Doelstelling Organisatie - Kwaliteitseisen
A4	24	I-2	Als beleid met kwalitatief onvoldoende data werkt moeten zij zich realiseren dat het eindresultaat ook kwalitatief tekort komt	Doelstelling Organisatie - Kwaliteitseisen
A4	25	I-4	Er moet onderscheid gemaakt worden in kwaliteitsaspecten en kwaliteitseisen die aan deze aspecten gesteld worden.	Kwaliteitseisen; Kwaliteitscriteria
A4	26	I-4	Als kwaliteitsaspecten kunnen o.a. onderkend worden: oVolledigheid – zijn objecten/gegevens die verplicht zijn ook daadwerkelijk geleverd of heb je iets aan de data als het niet verplicht is en ook niet is aangeleverd oGeometrische kwaliteit – zijn de geometrische gegevens in lijn met de realiteit in het veld (bijvoorbeeld locaties van objecten) oJuistheid – zijn de gegevens consistent met elkaar (bijvoorbeeld twee gekoppelde datasets die niet consistent zijn met elkaar) oActualiteit – hoe actueel zijn de gegevens oFormaat – In welk formaat moeten gegevens aangeleverd worden willen deze gegevens verwerkt kunnen worden	Kwaliteitscriteria
A4	27	I-5	Doelstellingen DQM komen tot uitdrukking in een GLP	Doelstelling DQM - Kwaliteitseisen

Figuur 23: Resultaten samenschap DQM met BDA - deel 1



Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
A5	28	I-2	Met traditionele data kun je denken aan criteria als volledigheid, tijdigheid en integriteit	Kwaliteitscriteria
A5	29	I-4	Kwaliteitseisen zijn afhankelijk van het beleidsvraagstuk wat speelt en moeten aangedragen worden vanuit de beleidsclusters en kunnen verdeeld worden naar: oTechnische eisen – Te controleren door gegevensmanagement oGebruikerseisen – Te controleren door de inhoudelijke specialisten van de beleidsclusters	Kwaliteitseisen
A5	30	I-4	Kwaliteitseisen komen tot uitdrukking in: oContract met de gegevensleverancier (voorwaarden waaraan de gegevensleverancier zich moet houden) oVerwachtingen die de gebruikers hebben ten aanzien van de kwaliteit (incl. inzicht in hoeverre de data voldoet aan de gestelde kwaliteitseisen)	Kwaliteitseisen
A6	31	I-1	Gegeven de eigenschappen van Big Data, uitgedrukt in de 5 V's, kan het zo zijn dat we de verwerking hiervan moeten onderbrengen in de Cloud (bijvoorbeeld dat volume dusdanig groot is dat we dit zelf niet kunnen opslaan en/of verwerken).	Eigenschappen BD - CC
A6	32	I-1	Variety: We hebben van doen met: a)Gestructureerde data b)Semi-gestructureerde data c)Ongestructureerde data	Eigenschappen BD
A6	33	I-2	Tijdigheid, actualiteit en frequentie/omloopsnelheid zijn belangrijke aspecten	Eigenschappen BD
A6	34	I-2	Daarnaast krijg je te maken verschillende bronnen met uiteenlopende formaten	Eigenschappen BD
A6	35	I-2	We moeten hier nog expertise in opbouwen	Expertise BDA
A6	36	I-4	In het kader van een opleiding eens gewerkt met een extract van een Big Data set	Expertise BDA
A6	37	I-4	Big Data stelt eisen aan de beschikbare middelen en die heb je niet altijd beschikbaar vandaar met een extract gewerkt.	Eigenschappen BD
A6	38	I-4	Bij Big Data waarbij bijvoorbeeld opnames zijn gemaakt zien we items terug die verschillende kenmerken bevatten zoals bijv. Tijdstip, locatie, kleur etc.	Eigenschappen BD
A6	39	I-4	Je kunt daarbij onderscheid maken in data die real-time wordt aangeleverd of data die als bestand beschikbaar is en opgeslagen wordt.	Eigenschappen BD
A7	40	I-1	Doordat we gegevens dan buiten de organisatie leggen moeten we de kwaliteitsdoelstellingen strenger bewaken.	Kwaliteitseisen - CC
A7	41	I-1	Doordat we minder mogelijkheden hebben om te controleren als de data zich buiten de organisatie bevindt, is het een uitdaging hoe we dit proces vormgeven.	Toetsen Datakwaliteit - CC
A7	42	I-1	Ook de verklaarbaarheid van wijzigingen in de data wordt dan een uitdaging daar we niet volledig grip hebben op het gehele proces	Kwaliteitseisen - CC
A7	43	I-1	Volume: Door hoge volumes kan er risico ontstaan op ontbrekende data/onvolledige datasets doordat bijvoorbeeld het grote volume niet meer in het systeem en/of database past.	Eigenschappen BD - Kwaliteitseisen
A7	44	I-1	Veracity: Je hebt met verschillende bronnen te maken die verschillende kwaliteitsniveaus hanteren waardoor het in samenhang kijken naar datasets vanuit de verschillende bronnen een probleem gaat vormen Bijvoorbeeld vanuit cluster wegbeheer kijken ze naar bomen en ook vanuit cluster natuur. Zijn die twee verzamelingen wel tegen elkaar af te zetten	Eigenschappen BD - Kwaliteitseisen
A7	45	I-1	Velocity: Dataverloop kan zodanig snel zijn dat het kwaliteitsproces dit niet kan bijbenen en dat je als het ware achter de feiten aanholt	Eigenschappen BD - Kwaliteitseisen
A7	46	I-1	Value: Kijkend naar de bij de inleiding geschetste piramide betekent geen goede kwaliteit van data in de basis dat dit doorwerkt in alle lagen van de piramide en uiteindelijk ook kwaliteit van de action/beslissing waardoor de waarde voor de organisatie onder druk komt te staan. Bewustzijn m.b.t. de kwaliteit van data moet op alle niveaus van de lagen van de piramide aanwezig zijn.	Doelstelling Organisatie - Kwaliteitseisen
A7	47	I-2	Beleid is richtinggevend	Doelstelling Organisatie - Kwaliteitseisen
A7	48	I-2	Kijkend naar een tabel kun je zowel qua aantal records (volledigheid) als inhoud van de records zelf kwaliteitseisen stellen waarbij je ook nog eens de mate van detail/granulariteit hierin kunt betrekken.	Kwaliteitseisen
A7	49	I-2	Actualiteitseis krijgt een prominentere plek	Kwaliteitseisen
A7	50	I-4	Je kunt de items en bijbehorende de kenmerken vergelijken met records en attributen van gewone data waarop je soortgelijke kwaliteitsaspecten kunt los laten als die gelen voor gewone data. Voorbeeld: je kunt een sensor hebben waar een beschikbaarheidsprobleem speelt waardoor je een issue hebt omtrent het aspect volledigheid (volledigheid heb je ook bij gewone data) of de sensor is op hol geslagen waardoor de juistheid van de data in twijfel kan worden getrokken (juistheid speelt ok bij gewone data)	Eigenschappen BD - Toetsing Datakwaliteit
A7	51	I-4	Gegeven de hoeveelheid van gegevens die bij Big Data speelt ligt er vaak een statistisch vraagstuk achter (bijvoorbeeld onderkennen van patronen).	Eigenschappen BD - Toetsen Datakwaliteit
A7	52	I-4	Om dit statistisch vraagstuk op te lossen moet sprake zijn van representatieve set om gerichte uitspraken te doen. Bijvoorbeeld het onderkennen van frauduleuze transacties t.o.v. de totale set aan transacties. Gegeven de kleine hoeveelheid frauduleuze transacties moet de totale set flink groot zijn.	Eigenschappen BD - Toetsen Datakwaliteit
A7	53	I-4	De omvang van de representatieve set is afhankelijk van het aantal kenmerken waarover je een uitspraak wilt doen ( een vraagstuk waarbij je maar over 1 kenmerk een uitspraak wilt doen kan een kleinere set zijn dan een vraagstuk waarbij je over 10 kenmerken een uitspraak wilt doen)	Eigenschappen BD - Toetsen Datakwaliteit

Figuur 24: Resultaten samenschap DQM met BDA - deel 2

Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
A8	54	I-1	Percentage goede data moeten meestal liggen tussen de 90 en 100% Afhankelijk van de impact van het kwaliteitsniveau op de besluitvorming moeten hogere eisen gesteld worden aan de data. Dus dit betekent dat je qua percentage goede data opschuift naar 100.	Doelstelling Organisatie - Toleranties
A8	55	I-1	Bij fouten moet onderscheid gemaakt worden in fouten o.b.v. toeval of structurele oorzaken. Structurele oorzaken moeten opgelost worden	Verbeteren Datakwaliteit
A8	56	I-1	Het begin bij goed nadenken over datgene wat wel/niet acceptabel is om welke reden. Welke afwijking wordt bijvoorbeeld geaccepteerd en wanneer is sprake van outliers.	Toleranties
A8	57	I-4	Toe te passen toleranties zijn afhankelijk van de eisen van de gebruiker	Doelstelling Organisatie - Toleranties
A8	58	I-5	We kennen geen toleranties wat betreft uit te voeren controles. Als het iets fout is gaat het terug	Toleranties
A9	59	I-1	Checks op gestructureerde data doen we al jaren. Denk aan ontbrekende waarde/null-values, invalid values en transcription errors (bijv. in getallen is ' ' in Engelse context een ' ' in Nederlandse context). Daarnaast kun je controles doen op toegestane bandbreedtes van waarden (min-max values), maximale percentages voor afwijkingen etc.	Toetsen Datakwaliteit
A9	60	I-1	Check op semigestructureerde data wordt al iets moeilijker. Denk aan een naam. Naam qua formaat kan juist zijn maar is het wel de juiste naam (bijv. juist gespeld). Dan moet je al meerdere waarden van Naam voor dezelfde persoon gaan vergelijken.	Toetsen Datakwaliteit
A9	61	I-1	Om controles op ongestructureerde data uit te kunnen voeren moet je de ongestructureerde data omzetten naar gestructureerde of semi-gestructureerde data zodat je vervolgens de technieken kunt toepassen die gelden bij kwaliteitschecks voor gestructureerde en semi-gestructureerde data gelden. Daarnaast kun je data groeperen/clusteren o.b.v. eigenschappen en van daaruit controles uitvoeren	Toetsen Datakwaliteit
A9	62	I-2	Toetsing van datakwaliteit op dit moment niet gestructureerd en incident gedreven en afhankelijk van de beschikbare capaciteit Liefst doen we dit automatisch maar voor BDA is dit nog best lastig maar wel belangrijk	Toetsen Datakwaliteit
A9	63	I-2	We zijn in veel geval geen eigenaar van de data (kopen data in). Dit betekent dat we bij fouten dit terugleggen bij de bron Uitzondering zijn landmetingen die we zelf doen	Verbeteren Datakwaliteit
A9	64	I-4	Je kunt op 3 manieren de data toetsen: oKijken naar de data zonder inhoudelijke kennis over de data (bijvoorbeeld consistenties tussen attributen of het aanwezig zijn van verplichte attributen) oProjectie van de data op de realiteit in het veld (door landmeters of medewerkers van de beleidsclusters) oKijken naar de data met inhoudelijke kennis waarbij je uitspraken doet over het realiteitsgehalte van de data	Toetsen datakwaliteit
A9	65	I-4	In geval je grote hoeveelheden data hebt, moet je kijken hoe je deze data gaat controleren. Kun je wel uit de voeten met een steekproef als er grote hoeveelheden data zijn?	Eigenschappen BD - Toetsen Datakwaliteit
A9	66	I-4	•Indien je de data niet zelf inwint maar er wel afspraken liggen ten aanzien van de kwaliteit van de gegevenslevering (bijv. bij CBS): In geval van fouten gaat de foutieve aanlevering terug naar de leverancier •Indien je de data niet zelf inwint en er geen afspraken liggen ten aanzien van de kwaliteit (bijv. in het geval van open data): In dat geval moet je vooraf een intake doen op de kwaliteit en vaststellen of het zinvol is deze data in te zetten (je hebt immers geen invloed op de kwaliteit) •Indien je de data zelf inwint: In dat geval ben je zelf verantwoordelijk voor de kwaliteit en bijbehorende kwaliteitsverbeterende acties. Het is dan van belang de oorzaak na te gaan en na te gaan of hier een structureel probleem achter ligt. Bijvoorbeeld een foutieve typering van Geo objecten waardoor je bij de GIS analyses in de problemen komt. Snelle oplossing is desbetreffende gevallen eruit filteren terwijl de eigenlijke oplossing is om dit in de bron op te lossen waarbij er nog mogelijk een achterliggend probleem speelt dat iemand dit structureel verkeerd typeert	Verbeteren Datakwaliteit
A9	67	I-5	We voeren volgende soorten controles uit: oOf data in overeenstemming is met de metadata oOf kolommen op een juiste wijze gevuld zijn oOf er geen lege rijen tussen zitten of rijen met ontbrekende kolommen oOf een beschrijving is toegevoegd waaraan staat aangegeven wat onder een kolom verstaan wordt oJuiste encoding toegepast oConsistent dimensies worden toegepast bij de diverse kolommen oControleren of geleverde data overeenkomt met het bovenliggend datatype waarbij je ook kijkt naar het waardebereik	Toetsen Datakwaliteit
A9	68	I-5	De gegevens in de database zijn afhankelijk van de gegevensleveringen. Dus als je de leveringen controleert of zij voldoen aan het GLP en aanvullende controles heb je ook impliciet de gegevens gecontroleerd zoals deze in de database landen	Toetsen Datakwaliteit
A9	69	I-5	Bepaalde data moet gewoon 100% goed zijn zonder dubbele rijen. Als hier een fout in zit vragen we een herlevering	Verbeteren Datakwaliteit
A9	70	I-5	Wij moeten de controles wel binnen 30 dagen na levering uitgevoerd hebben	Toetsen Datakwaliteit
A9	71	I-5	Kleine dingen die we sneller zelf hebben opgelost dan opgelost worden door een herlevering dan doen we dit zelf incidenteel. We passen de data zelf niet aan. Het betreft hier zaken als spellingsfouten in kolomnamen etc. De zaken die zeer duidelijk zijn, en zonder aannames kunnen worden uitgevoerd	Verbeteren Datakwaliteit
A9	72	I-5	We willen echter naar een situatie toe waarbij we zelf zo min mogelijk hoeven te controleren en vragen aan de leveranciers deze controles zelf uit te voeren aan de hand van GLP alvorens de data aan ons te leveren	Verantwoordelijkheden
A9	73	I-5	Een belangrijk punt vorm de verantwoordelijkheden: waar ligt wel verantwoordelijkheid. Momenteel voeren we hier discussies over. Wat moeten we zelf en wat is de verantwoordelijkheid van de gegevensleverancier	Verantwoordelijkheden
A9	74	I-5	oWe hebben technische controles die worden uitgevoerd door Gegevensmanagement oWe hebben inhoudelijke validaties (komt de data overeen met de realiteit). Deze worden uitgevoerd door de beleidsclusters	Toetsen Datakwaliteit
A9	75	I-5	We zijn bezig om controles, verbeteringen en terugkoppeling te optimaliseren uitgaande van verschillende scenario's, rekening houdend met capaciteit en middelen	Toetsen Datakwaliteit

Figuur 25: Resultaten samenhang BDA met DQM - deel 3

### *Samenvatting analyse resultaten*

Onderstaande tabel geeft een samenvatting van de analyse resultaten waarbij, over de tekstfragmenten heen, het aantal keer is geteld dat een term of relatie tussen termen is voorgekomen. Het aantal weerspiegelt daarbij de mate van herkenbaarheid van een term of relatie tussen termen.

Term of relatie tussen termen	Aantal
Doelstelling Organisatie	3
Doelstelling Organisatie - Doelstelling DQM	2
Doelstelling Organisatie - Kwaliteitseisen	9
Doelstelling Organisatie - Doelstelling BDA	3
Doelstelling Organisatie - Toleranties	2
Doelstelling DQM	3
Doelstelling DQM - Kwaliteitseisen	1
Doelstelling BDA	4
Kwaliteitseisen	6
Kwaliteitseisen - Cloud Computing	2
Kwaliteitscriteria	3
Eigenschappen Big Data	6
Eigenschappen Big Data - Kwaliteitseisen	3
Eigenschappen Big Data - Toetsen Datakwaliteit	5
Eigenschappen Big Data - Cloud Computing	1
Toleranties	2
Toetsen Datakwaliteit	12
Toetsen Datakwaliteit - Cloud Computing	1
Verbeteren Datakwaliteit	5
Verantwoordelijkheden	2
Expertise BDA	2

*Figuur 26: Samenvatting analyse resultaten samenhang DQM met BDA*



## Bijlage III: Resultaten wijze waarop DQM zich manifesteert in een EA

Deze bijlage geeft voor de wijze waarop DQM zich manifesteert in een EA inzicht in:

- Onderkende termen vanuit de theoretische verkenning
- Geformuleerde vragen die als leidraad golden voor de gevoerde interviews
- Resultaten uit de gevoerde interviews en daaraan gekoppelde analyse
- Samenvatting van de analyse resultaten

### Onderkende termen

O.b.v. de verkregen inzichten vanuit een theoretisch oogpunt zoals weergegeven in sectie 4.3.1 zijn de volgende termen te onderkennen.

Termen
Doelstelling Organisatie
DQ Beleid
DQ Standaarden
Maatregelen Datakwaliteit (DQ)
Business Laag
Organisatie
Processen
Information Systems Laag
Applicaties
Data
Infrastructuur Laag
Infrastructuur
Data Integratie

Figuur 27: Termen DQM uitgedrukt in een EA

### Geformuleerde vragen

Op basis van het verkregen theoretisch inzicht en daaraan gekoppelde termen zijn onderstaande vragen geformuleerd.

Nr. Vraag	Vraag
B1	Op welke wijze wordt binnen de organisatie gebruik gemaakt van een EA en op welke wijze wordt dit beheerd/actueel gehouden
B2	Op welke samenhangende wijze is DQM tot uitdrukking gebracht in de verschillende lagen (business, information systems en infrastructuur) van een EA met aandacht voor: <ul style="list-style-type: none"><li>•Doelstellingen voor DQM</li><li>•DQ beleid en DQ standaarden</li><li>•Organisatie en processen</li><li>•Applicaties en Data</li><li>•Infrastructuur</li></ul>
B3	In hoeverre is bij maatregelen voor borging van datakwaliteit onderscheid te maken in databeheer en datagebruik met bijbehorende verantwoordelijkheden
B4	In hoeverre is bij maatregelen voor borging van datakwaliteit onderscheid te maken in toetsing van datakwaliteit en verbetering (corrigerende acties) voor data zodat beantwoord wordt aan de gestelde kwaliteitseisen

Figuur 28: Vragen DQM uitgedrukt in een EA

### Resultaten uitgevoerde interviews en daaraan gekoppelde analyses

In onderstaande tabel zijn de antwoorden op de gestelde vragen weergegeven waarbij o.b.v. de tekst is gekeken aan welk term of relatie van termen dit te relateren is. Een antwoord kan hierbij opgeknipt zijn in meerdere tekstfragmenten om het vervolgens te kunnen relateren aan de termen en onderlinge relaties. Een groot deel van de termen en onderlinge relaties hebben hun oorsprong in het verkregen theoretisch inzicht. Een deel van de termen en onderlinge relaties kunnen aanvullend geïntroduceerd zijn n.a.v. de gegeven antwoorden.

Toelichting kolommen uit de tabel:

- Nr. Vraag: Link naar de tabel met geformuleerde vragen (het nummer correspondeert met het nummer in de tabel met vragen)
- Nr. Antwoord: Identificatie van een tekstfragment van het antwoord op een vraag
- Interview: Identificatie van het interview
- Tekst: Tekstfragment van een antwoord wat te koppelen is aan een term of relatie tussen termen
- Term of relatie tussen termen: Termen of relatie tussen termen die herkend kunnen worden vanuit het tekstfragment (bij relatie tussen termen zijn de termen gescheiden door een “-”)

Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
B1	1	I-2	We hebben een referentie architectuur (soll positie) en er is een beeld van de ist.	Adoptie EA
B1	2	I-3	Er is sprake van een overheidsbrede referentiearchitectuur die actueel wordt gehouden. Doorvertaling naar de architectuur van de organisatie moet geactualiseerd worden	Adoptie EA
B1	3	I-3	Bij aanbestedingen moeten leveranciers standaarden van het forum (vanuit overheidswege) voor standaardisatie respecteren (zoals standaarden m.b.t. datacommunicatie, gebruik StUF etc., xml maar ook standaarden die te maken hebben met data).	Standaarden Algemeen
B1	4	I-5	Vanuit Gegevensmanagement kijkend naar een EA zitten we met name op de information systems laag (bijv. tooling en hoe zetten we dit/bieden we dit aan) Wat betreft de Infrastructuur zijn infrastructuur mensen aan zet., in samenwerking met diverse personen van team gm. Ten aanzien van de business laag houden we ons bezig met producten (dashboards, viewers) en de processen om vanuit de bron te komen tot deze producten, rekening houdend met de waarde voor de organisatie	Adoptie EA
B1	5	I-8	Bij introductie van een nieuwe informatievoorziening is er sprake van architectuur check	Adoptie EAM
B1	6	I-8	We werken met een EA met 9 vlakken en 2 aanvullende vlakken voor informatiebeveiliging en beheer.	Adoptie EA
B1	7	I-8	Als een verandering speelt wordt per vlak de impact uitgedrukt in een PSA (Project Start Architectuur). Je kunt het zien als een soort checklist.	Adoptie EAM
B1	8	I-8	Voor elk vlak zijn principes beschreven.	Adoptie EAM
B1	9	I-8	We werken hierbij volgens de DYA methodiek waarbij de regel is: pas principe toe of leg uit waarom je van een principe afwijkt met motivatie (waarbij er de mogelijkheid is dat je op langere termijn alsnog aan het principe voldoet)	Adoptie EAM
B1	10	I-8	Het is niet de bedoeling om als een politieagent op te treden. Er moeten een bepaalde mate van flexibiliteit zijn anders is het niet werkbaar. We streven wel zoveel mogelijk naar standaardisatie.	Adoptie EAM
B1	11	I-8	We hebben een EA en hebben een PSA per project.	Adoptie EA; Adoptie EAM
B1	12	I-8	Er sprake van een architectuur-team waar EA architecten en informatiearchitecten in zitting hebben en die de projecten beoordelen/bewaken	Adoptie EAM
B2	13	I-2	Vanuit Gegevensmanagement is er echter geen architectuurplaat	Adoptie EA voor DQM
B2	14	I-2	We zouden moeten toewerken naar gebruik van een EA zoals we dit doen in het kader van een hitte kaart voor risk en security management Je zou iets soortgelijks moeten krijgen van DQM: wat heeft impact wat vormt een risico. Het heeft potentieel echter er ontbreekt een EA en een DG beleid ontbreekt	Adoptie EA voor DQM
B2	15	I-2	Er zijn geen DQ beleid en standaarden. We werken momenteel Bottom up	DQ Beleid en DQ Standaarden
B2	16	I-2	We zouden daar een weergave van een IST moeten komen die we toetsen om vervolgens tot een SOLL te komen met de noodzakelijke verbeteringen. Denk aan het gebruik van Archimate i.k.v. risl/security management waarbij pijnpunten inzichtelijk worden gemaakt. Voor DQM is dat nog een onontgonnen gebied	Adoptie EA voor DQM
B2	17	I-3	DQM is niet tot uitdrukking gebracht in een EA. Nog nooit eerder tegen gekomen	Adoptie EA voor DQM
B2	18	I-3	EA is ook te hoog over om hier maatregelen in het kader van DQM in tot uitdrukking te brengen	Adoptie EA voor DQM
B2	19	I-4	Geen zicht op de doelstellingen voor DQM en de vertaling hiervan naar beleid en standaarden m.u.v. algemene architectuurprincipes die gelden.	Doelstelling DQM - DQ Beleid en DQ Standaarden
B2	20	I-4	Data Kwaliteit is voor Gegevensmanagement geen doel op zich en vormt eigenlijk een verantwoordelijkheid van de beleidsvelden	Doelstelling Organisatie - Doelstelling DQM
B2	21	I-4	Eisen ten aanzien van datakwaliteit zijn verschillend en zijn afhankelijk van het doel waarvoor de data wordt ingezet dus je kunt dit niet generaliseren	Doelstelling Organisatie - Datakwaliteit
B2	22	I-4	Je zou kunnen toewerken naar een standaardisatie in de vorm van een checklist voor datakwaliteit o.b.v. het datavraagstuk wat speelt	DQ Beleid en DQ Standaarden
B2	23	I-4	Je zou kunnen kijken naar de diverse kwaliteitsaspecten die een rol kunnen spelen en vervolgens kunnen kijken hoe en waar het toetsen belegd. Doe je dit manueel of leg je het in de automatisering. Nu vindt het veelal manueel plaatsen gegeven de verschillende datasets, verschillende afnemers en de verschillende kwaliteitsaspecten die spelen.	Datakwaliteit - Inrichting Toetsen
B2	24	I-4	Los van of je dit manueel of geautomatiseerd doet is er veel tijd/capaciteit en geld mee gemoeid. Dat betekent dat een kosten/baten vraagstuk hier om de hoek komt kijken. Kijkend naar de huidige situatie is de vraag of het geautomatiseerd uitvoeren van toetsen wel een kostenvoordeel oplevert.	Datakwaliteit - Inrichting Toetsen

Figuur 29: Resultaten DQM uitgedrukt in een EA – deel 1



Nr. Vraag	Nr. Antwoord	Interviu	Tekst	Term of Relatie tussen termen
B2	25	I-4	Als je kijkt naar Big Data ontken je er niet aan om het in de automatisering te leggen. Je krijgt een omslag van handmatig valideren naar geautomatiseerd valideren	BD - Datakwaliteit - Inrichting Toetsen
B2	26	I-4	In geval van Big Data moet je de toetsing veelal geautomatiseerd gaan inrichten en dat betekent dat je vooraf moet nadenken over kwaliteitsaspecten en bijbehorende eisen en hoe je dit wilt toetsen. Wat dat betreft vormt Big Data een drijfveer om dit structureel goed in te richten. Nu vindt het toetsen manueel plaats o.b.v. een aangeleverde dataset en heeft veelal een ad-hoc karakter.	BD - Datakwaliteit - Inrichting Toetsen
B2	27	I-4	Introductie van Big Data en de noodzaak om vooraf goed na te denken over datakwaliteit en het geautomatiseerd inrichten van noodzakelijke controles kan bijdragen in het creëren van bewustzijn rondom datakwaliteit zowel aan de gegevensverwerkende kant als de gebruikers kant	BD - Datakwaliteit - Inrichting Toetsen
B2	28	I-4	Als je kijkt naar de verantwoordelijkheden voor het toetsen van de kwaliteit kun je concluderen dit op dit moment niet goed belegd is en zou je moeten overgaan naar een goede rolverdeling tussen datgene wat de informatiespecialist controleert (technisch gericht) en datgene wat de beleidsmedewerker controleert (inhoudelijk gericht)	Verantwoordelijkheden
B2	29	I-5	We hebben afspraken met gegevensleveranciers die vastgelegd zijn in een GLP en hoe kunnen we dit verbeteren	DQ Beleid en DQ Standaarden
B2	30	I-5	We kijken naar verantwoordelijkheden in dat kader. Ook kijkend naar beheer. Soms ligt dit nu bij de informatiemakelaar waar dit niet thuis hoort. Dit moeten we overdragen naar de informatiespecialisten B zodat de informatiemakelaar zich meer kan bezighouden met de kwaliteit van de data, regels en trends en niet met beheertaken.	Verantwoordelijkheden
B2	31	I-5	Ten aanzien van datakwaliteitsstandaarden (bijv. meta standaarden). We kunnen datakwaliteitseisen nog meer in standaarden gieten.	DQ Beleid en DQ Standaarden
B2	32	I-5	Ten aanzien van datakwaliteitsbeleid wordt dit op dit moment met name bottom-up vormgegeven en dat zou meer top-down moeten.	DQ Beleid en DQ Standaarden
B2	33	I-5	We zijn wel stappen aan het maken, kijkend naar het manifest "zo werken we met data" wat in ontwikkeling is	DQ Beleid en DQ Standaarden
B2	34	I-5	Ten aanzien van de processen voor DQM: oProcessen zijn nog veel informeel ingericht oZouden moeten we meer structuur in hierin moeten aanbrengen tevens rekening houdend met de rolverdeling om te voldoen aan de vraagstelling vanuit beleid oHoe gaan we het GLP verder inrichten vertaald naar interne processen oWe zijn met verbeteringen/optimalisering bezig o.b.v. scenario's.	Processen
B2	35	I-5	We werken volgens een ETL werkwijze waarbij data inwinnen, transformeren (in een andere structuur gieten) en importeren in de toepassingen (zoals ABF swing).	Processen
B2	36	I-5	Voor controle op de data worden scripts gemaakt. Deze voeren we uit voor de transformatieslag.	IT systemen
B2	37	I-5	Daarnaast moet je kijken naar instellingen van de toepassingen zelf	IT systemen
B2	38	I-5	Ook hier zijn we bezig met verbeterlagen waarbij er een uitwisseling speelt tussen het proces en daarbij gebruikte toepassingen rekening houdend met het principe data bij de bron	Processen - IT systemen
B2	39	I-5	Daarnaast zijn we bezig met reductie van het applicatie landschap rekening houdend met overlappende functionaliteiten.	IT Systemen
B2	40	I-5	We moet rekening houden met flink aantal datasets. Daar gaat flink wat tijd in zitten. Gebruik maken van scripts levert hier veel tijds winst	IT Systemen
B2	41	I-5	Op dit moment zitten die checks laat in het proces en reeds voorbij de 30 dagen termijn. Dat wil je meer naar voren trekken.	Processen
B2	42	I-5	Verbeterslagen gaan geleidelijk aan (kleine stappen zetten). Want winkel moet tussentijds open blijven maar we willen de winkel efficiënter maken	Verbetering inrichting
B2	43	I-5	We moeten de verbeteringen vooruitstrevend en proactief benaderen waarbij de grenzen moeten opzoeken met aandacht voor de medewerkers	Verbetering inrichting
B2	44	I-5	We moeten daarbij breder kijken naar alleen beheer. We moeten he niet te zeer vanuit de eigen koker bekijken. Dat geldt overigens niet voor iedereen	Integratie
B2	45	I-6	Doelstelling is om de basisbestanden op orde te hebben. Zij vormen immers basis voor verdere data-analyses	Doelstelling DQM
B2	46	I-6	Er kan onderscheid worden gemaakt in technische controles (verantwoordelijkheid Gegevensmanagement) en inhoudelijke controles (verantwoordelijkheid beleidsmedewerker)	Verantwoordelijkheden
B2	47	I-6	We zien echter dat de beleidsmedewerker niet altijd volledig zijn rol pakt bij het uitvoeren van de inhoudelijke controles. De verantwoordelijkheden kunnen duidelijker belegd worden.	Verantwoordelijkheden
B2	48	I-6	Ten aanzien van standaarden kan aangegeven worden dat er ten aanzien van de databestanden geen standaarden zijn. Voor de metadata gekoppeld aan databestanden is dat wel het geval. Deze worden veelal opgesteld door een externe organisatie	DQ Beleid en DQ Standaarden
B2	49	I-6	Kwaliteitseisen waarop getoetst wordt zijn: oVerplichte attributen oGeometrieën: geen gekke dingen (bijv. kruisende lijnen); geen overlappende objecten binnen 1 laag en geen overlappende objecten over lagen heen (voor zover niet toegestaan) oAttributen/records behorende bij een Geo object in lijn met de geometrie oGeen Multipolygonen (1 record met meerdere geometrieën)	Kwaliteitseisen

Figuur 30: Resultaten DQM uitgedrukt in een EA - deel 2

Nr. Vraag	Nr. Antwoord	Interviu	Tekst	Term of Relatie tussen termen
B2	50	I-8	Ten aanzien van data hanteren we standaarden zoals data bij de bron en enkelvoudig beheer – meervoudig gebruik. Zodat je ook de datakwaliteit kan borgen en niet dat er meerdere waarheden gaan ontstaan.	DQ Beleid en DQ Standaarden
B2	51	I-8	Voorkomen dat het op lange termijn problemen ontstaan. We moeten ervoor zorgen dat er geen meerdere systemen gaan ontstaan die dezelfde doelstelling/functionaliiteit inhouden	DQ Beleid en DQ Standaarden
B2	52	I-8	We zijn als afdeling voor organisatie en informatie veelal geen eigenaar van de data Eigenaarschap ligt bij andere afdelingen . Wij beheren, faciliteren en adviseren	Databeheer
B2	53	I-8	Eigenaar is verantwoordelijk voor de kwaliteit van de data. Verantwoordelijke afdelingen moeten dit organiseren. Afdeling voor organisatie en informatie heeft kaders en richtlijnen maar kunnen geen grote invloed uitoefenen op de datakwaliteit.	Verantwoordelijkheden
B2	54	I-8	Het is niet bij alle afdelingen goed georganiseerd waardoor je verschillen in datakwaliteit ziet. Het is een gebied wat verder ontwikkeld moet worden middels een datafundament en manifest “zo werken we met data”	Verantwoordelijkheden; DQ Beleid en DQ Standaarden
B2	55	I-8	Er zijn kaders en richtlijnen voor het aanschaffen, inwinnen, bewerken en visualiseren van data. Daar adviseren we ook in. Het kan echter wat strakker georganiseerd worden	DQ Beleid en DQ Standaarden
B2	56	I-8	Het Data Quality Management is op dit moment nog niet op eenduidige manier ingevuld. Het verschilt per afdeling.	DQ Beleid en DQ Standaarden
B2	57	I-8	Daarnaast zijn we aangehaakt op landelijke voorzieningen. Daar zijn ook kaders aan gekoppeld. Ketensamenwerking, kernregistraties en basisregistraties inclusief daarbij horende standaarden spelen daar een rol.	DQ Beleid en DQ Standaarden
B2	58	I-8	Bij aanschaf van data wordt ook een document opgesteld met afspraken rondom het verwachte kwaliteitsniveau van de data.	DQ Beleid en DQ Standaarden
B2	59	I-8	De crux is om hier generieke afspraken te maken en hier een standaardisatie in door te voeren en af te stappen van ad hoc methoden. Dit zal ook bevorderlijk werken ten aanzien van de datakwaliteit	DQ Beleid en DQ Standaarden
B2	60	I-8	Het is complexe materie rekening houdend met ketens en variatie aan bronnen die spelen waarbij de verschillende sectoren dit ook nog eens allemaal anders benaderen.	Ketensamenwerking
B3	61	I-2	Databeheer: voor sommige data voeren vanuit een wettelijke taak het beheer uit denk aan BGT waaraan ook kwaliteitseisen aan verbonden zijn	Databeheer
B3	62	I-2	Voor datagebruik is er voor gegevensmanagement een indirecte relatie met de primaire processen	Datagebruik
B3	63	I-3	Er gelden algemene regels als eenmalige opslag en meervoudig gebruik en algemene consistentieregels maar verder gaat dit niet	Databeheer; Datagebruik
B3	64	I-3	Algemeen principe voor eenmalige opslag een meervoudig gebruik	Databeheer; Datagebruik
B3	65	I-3	Bij datagebruik eisen we dat data uit de bronsystemen wordt opgehaald en geen kopieën getrokken worden	Datagebruik
B3	66	I-3	Geldt ook voor de basisregistraties (KvK, NHR, BAG etc.). Applicaties moeten data hieruit halen middels gebruik maken van services voor actuele bevragingen. Dit is de praktische uitwerking van het algemeen principe	Datagebruik
B3	67	I-3	Dit algemeen principe is ook een vertaling van het NORA principe waarin je de burger eenmalig vraagt naar zijn gegevens	Datagebruik
B3	68	I-5	Bij fouten gaan we terug naar de bron. Daar ligt het beheer	Databeheer
B3	69	I-8	In de beheerfase, daar waar geen sprake is van een eenmalige levering, moet je die kwaliteit blijven garanderen en dat is best een lastig verhaal. Dat mechanisme is nu beperkt tenzij we de data zelf beheren want dan hebben we er wel kijk op. Dat wil niet zeggen dat het bij externe leveringen altijd fout gaat. Het gaat in sommige gevallen ook wel goed. Licht ook aan de leverancier van de gegevens en met budget wat je hiervoor beschikbaar stelt.	Databeheer
B3	70	I-8	Je moet daarbij wel altijd het doel in het achterhoofd houden. Als je eenmalig een rapport wil maken o.b.v. een eenmalige levering is dat iets anders dan de leveringen die een repeterend karakter hebben.	Doelstelling Organisatie - Datagebruik
B3	71	I-8	Je zou ook kunnen toewerken naar een datawarehouse met data waarmee je ook toekomstige (latente) vraagstukken zou kunnen beantwoorden maar dat is gewoonweg niet mogelijk omdat je niet weet wat de vraag zou kunnen zijn.	Datagebruik
B3	72	I-8	Om data te verzamelen en te beheren heb je nu veel tijd en geld nodig. Je moet hier keuzes in maken. En meestal heeft men het geld niet beschikbaar om een grote dataset te beheren.	Databeheer; Datagebruik
B3	73	I-8	Het is belang hierin het eigenaarschap te onderkennen en bijbehorende verantwoordelijkheden ,in lijn hiermee, goed beleggen. Er wordt iets te snel naar Gegevensmanagement gekeken maar Gegevensmanagement moet dit terugleggen bij degene die daar verantwoordelijk voor is.	Verantwoordelijkheden
B4	74	I-3	Er zijn op dit moment, behalve de actualiteitseis, geen aanvullende kwaliteitsattributen onderkend en de wijze waarop dit is doorvertaald naar maatregelen voor borging van deze kwaliteit is niet tot uitdrukking gebracht in een EA	Maatregelen Borging Datakwaliteit
B4	75	I-3	Invulling geven aan de borging van de datakwaliteit laten we over aan de leverancier	Maatregelen Borging Datakwaliteit
B4	76	I-3	Dit is nog een ontwikkelingsgebied binnen onze organisatie	Maatregelen Borging Datakwaliteit
B4	77	I-3	Op dit moment zie ik als informatiearchitect/enterprise architect geen problemen op het gebied van datakwaliteit	Adoptie EA voor DQM
B4	78	I-3	Je zou een aantal niveaus moeten afdalen tot het niveau van gegevensmanagement. Daar zie ik wel aanknopingspunten	Maatregelen Borging Datakwaliteit
B4	79	I-4	Elke afnemer heeft ten aanzien van de verschillende kwaliteitsaspecten eigen kwaliteitseisen. Hoe gaat je dit vertalen naar de inrichting van je toetsen?	Doelstelling Organisatie - Datakwaliteit - Inrichting Toetsen

Figuur 31: Resultaten DQM uitgedrukt in een EA - deel 3



Nr. Vraag	Nr. Antwoord	Interviu	Tekst	Term of Relatie tussen termen
B4	80	I-4	Voor zover kwaliteitseisen in het verlengde van elkaar liggen zou je kunnen uitgaan van toetsing aan de strengste eis maar eisen hoeven niet altijd in het verlengde van elkaar te liggen	Datakwaliteit - Inrichting Toetsen
B4	81	I-6	Voordat we een databestand opnemen in Oracle voeren we de volgende checks uit: - Een topologische check m.b.v. ArcGis - Check of de metadata klopt	Toetsen Datakwaliteit
B4	82	I-6	Correctie geconstateerde fouten: oVoor zover we het databestand zelf hebben gemaakt (door bijvoorbeeld beleidsmedewerker) los we de fouten in samenspraak met de beleidsmedewerker manueel zelf op of laten we de beleidsmedewerker dit zelf oplossen (afhankelijk van het expertiseniveau van de beleidsmedewerker) oVoor zover we databestanden extern betrekken en hier specifieke opdrachten aan ten grondslag liggen, leggen we dit (via tussenkomst van de beleidsmedewerker) terug bij de externe partij, gevolgd door of een rechtstreekse afstemming tussen gegevensmanagement en de externe partij of het loopt via de beleidsmedewerker oVoor zover we databestanden extern betrekken en hier geen specifieke opdracht aan ten grondslag ligt koppelen we fouten terug voor zover hier afspraken over gemaakt zijn (bijv. kadaster)	Verbeteren Datakwaliteit - Verantwoordelijkheden
B4	83	I-6	Bij grote bestanden is de vraag hoe je de controle uitvoert: handmatig steekproefsgewijs of geautomatiseerd	Processen
B4	84	I-6	We zien dat bij het uitvoeren van controles veel van de kennis in de hoofden van de medewerkers zitten en de mate van nauwkeurigheid van de uit te voeren controles per persoon verschilt. Vroeger maakten we gebruik van checklisten waardoor de kennis meer geëxpliciteerd is en controles uniform kunnen worden uitgevoerd.	Processen
B4	85	I-6	Voor de meeste controles gebruiken we ArcGis als ondersteunend systeem	IT systemen

Figuur 32: Resultaten DQM uitgedrukt in een EA - deel 4

### Samenvatting analyse resultaten

Onderstaande tabel geeft een samenvatting van de analyse resultaten waarbij, over de tekstfragmenten heen, het aantal keer is geteld dat een term of relatie tussen termen is voorgekomen. Het aantal weerspiegelt daarbij de mate van herkenbaarheid van een term of relatie tussen termen.

Term of relatie tussen termen	Aantal
Adoptie EA	5
Adoptie EA voor DQM	6
Adoptie EAM	6
Doelstelling Organisatie - Datakwaliteit	2
Doelstelling Organisatie - Datakwaliteit - Inrichting Toetsen	1
Doelstelling Organisatie - Datagebruik	1
Doelstelling DQM	1
Doelstelling DQM - Beleid en Standaarden voor DQ	1
Beleid en Standaarden voor DQ	15
Standaarden Algemeen	1
Processen	5
Processen - IT systemen	1
Verantwoordelijkheden	7
Verbeteren Datakwaliteit - Verantwoordelijkheden	1
IT Systemen	5
Kwaliteitseisen	1
Big Data - Datakwaliteit - Inrichting Toetsen	3
Maatregelen Borgen Datakwaliteit	4
Datakwaliteit - Inrichting Toetsen	3
Toetsen Datakwaliteit	1
Data Beheer	7
Datagebruik	9
Verbetering Inrichting	2
Ketensamenwerking	1
Integratie	1

Figuur 33: Samenvatting analyse resultaten DQM uitgedrukt in een EA

## Bijlage IV: Resultaten wijze waarop CC voor BDA ingrijpt op een EA

Deze bijlage geeft voor de wijze waarop CC voor BDA ingrijpt op een EA inzicht in:

- Onderkende termen vanuit de theoretische verkenning
- Geformuleerde vragen die als leidraad gold voor de gevoerde interviews
- Resultaten uit de gevoerde interviews en daaraan gekoppelde analyse
- Samenvatting van de analyse resultaten

### Onderkende termen

O.b.v. de verkregen inzichten vanuit een theoretisch oogpunt zoals weergegeven in sectie 4.4.1 zijn de volgende termen te onderkennen.

Termen
Big Data Pipeline
Onderdelen Business Laag
Onderdelen Information Systems Laag
Onderdelen Infrastructure Laag
Doelstelling Organisatie
Cloud Service Model (CSM)
Cloud Deployment Model (CDM)
Impact Inrichting Business Laag
Impact Inrichting Information Systems Laag
Impact Inrichting Infrastructurele Laag

Figuur 34: Termen wijze waarop CC voor BDA ingrijpt op een EA

### Geformuleerde vragen

Op basis van het verkregen theoretisch inzicht en daaraan gekoppelde termen zijn onderstaande vragen geformuleerd.

Nr. Vraag	Vraag
C1	Welke stappen worden onderkend in een Big Data Analytics proces (zgn. Big Data Pipeline)
C2	Op welke wijze is de inrichting van deze stappen te projecteren op de onderdelen/lagen van een EA in het algemeen en de Business Laag in het bijzonder
C3	Op welke wijze is bij deze stappen de data kwaliteit geborgd en welke maatregelen zijn hiervoor in deze stappen geïntroduceerd
C4	Op welke wijze zijn deze maatregelen voor data kwaliteit te projecteren op de onderdelen/lagen van een EA in zijn algemeenheid en de Business laag in het bijzonder
C5	Welke doelstellingen heeft de organisatie geformuleerd die van invloed zijn op de keuze in het toe te passen Cloud Service Model (IAAS, PAAS, SAAS etc.) en Cloud Deployment Model (Public, Hybrid, Private)
C6	Heeft de organisatie reeds een keuze gemaakt inzake het toe te passen Cloud Service Model en Cloud Deployment Model, op basis waarvan is deze keuze gemaakt en welke criteria liggen hieraan ten grondslag
C7	Op welke wijze zijn deze criteria te koppelen aan de doelstellingen die de organisatie nastreeft in zijn algemeenheid en doelstellingen voor BDA in het bijzonder
C8	Op welke wijze ziet men dat de stappen binnen het BDA proces ondersteund door Cloud Computing vertaald naar de onderdelen van de EA
C9	Welke consequenties heeft dit voor de onderdelen/lagen van de EA voor de borging van de data kwaliteit in zijn algemeenheid en de Business laag in het bijzonder vanuit de doelstelling die voor DQM zijn geformuleerd en gerelateerde eisen voor data kwaliteit

Figuur 35: Vragen wijze waarop CC voor BDA ingrijpt op een EA

### Resultaten uitgevoerde interviews en daaraan gekoppelde analyses

In onderstaande tabel zijn de antwoorden op de gestelde vragen weergegeven waarbij o.b.v. de tekst is gekeken aan welk term of relatie van termen dit is te relateren. Een antwoord kan hierbij opgeknipt zijn in meerdere tekstfragmenten om het vervolgens te kunnen relateren aan de termen en onderlinge relaties. Een groot deel van de termen en onderlinge relaties hebben hun oorsprong in

het verkregen theoretisch inzicht. Een deel van de termen en onderlinge relaties kunnen aanvullend geïntroduceerd zijn n.a.v. de gegeven antwoorden.

Toelichting kolommen uit de tabel:

- Nr. Vraag: Link naar de tabel met geformuleerde vragen (het nummer correspondeert met het nummer uit de waarde bij de eerste kolom in de tabel met vragen)
- Nr. Antwoord: Identificatie van een tekstfragment van het antwoord op een vraag
- Interview: Identificatie van het interview
- Tekst: Tekstfragment van een antwoord wat te koppelen is aan een term of relatie tussen termen
- Term of relatie tussen termen: Termen of relatie tussen termen die herkend kunnen worden vanuit het tekstfragment (bij relatie tussen termen zijn de termen gescheiden door een “-”)

Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
C1	1	I-1	Stappen die onderkend worden in de voorbereiding van data op de weg naar informatie zijn: •Extract (E) •Transform (T) •Load (L) Het kan echter zo zijn dat we Transform en Load omwisselen De ETL vorm is de meer traditionele vorm de ELT vorm komt vaak voor in BD trajecten.	Big Data Pipeline
C1	2	I-1	Data-integratie is een belangrijke factor in al deze 3 stappen.	Data-Integratie
C1	3	I-2	Te onderkennen stappen: •Bron •Inwinnen/staging •Opslaan/staging •Analyses en komen tot tussenproduct •Verfijnen analyseresultaten om te komen tot uiteindelijke product •Visualisatie	Big Data Pipeline
C3	4	I-1	In alle bovengenoemde stappen kunnen zich kwaliteitsissues voor doen. Het is dus zaak in alle drie de stappen maatregelen te treffen om de kwaliteit te borgen. Bewerkingen kunnen namelijk leiden tot fouten maar ook de introductie van nieuwe bronnen zijn vaak in het begin oorzaak van kwaliteitsissues.	Datakwaliteit
C3	5	I-2	DQM heeft betrekking op alle stappen Inwinnen -> technische controle Analyse: zuiverheid, harde/zachte beleidsregels; Stamboom? Transparantie Resultaat voldoet aan gewenste visualisatie	Datakwaliteit
C5	6	I-1	Je moeten kijken naar het doel wat je voor ogen hebt als organisatie met de introductie van CC. Je moet het niet doen omdat iedereen het doet.	Doelstelling Organisatie - Toe te passen CSM/CDM
C5	7	I-2	Soll: vraag moet leidend zijn we werken echter bottom up	Doelstelling Organisatie - Doelstelling CC
C5	8	I-2	We zouden cloud moeten inzetten om grenzen te verleggen	Doelstelling Organisatie - Doelstelling CC
C5	9	I-2	Er is geen strategie voor CC. Als obstakels hebt zou je CC kunnen inzetten maar als genoeg Euro's beschikbaar zijn kun je ook je infrastructuur on premise aanpassen	Doelstelling Organisatie - Doelstelling CC
C5	10	I-2	Je zou als gegevensmanagement ontzorgd moeten worden	Doelstelling Organisatie - Doelstelling CC
C6	11	I-2	Realiteit is dat het nu nog incident gedreven is en dan cloud inzetten voor infrastructurele aangelegenheden als zich een probleem voordoet	Doelstelling CC
C6	12	I-2	We hebben geen cloudstrategie voor gegevensmanagement en er is geen koppeling met beleid	Doelstelling Organisatie - Doelstelling CC
C6	13	I-2	Voor nieuwe toepassingen wordt echter een cloud SaaS oplossing als voorwaarde gesteld terwijl dit voor gegevensmanagement nog niet aan de orde is.	Doelstelling Organisatie - Doelstelling CC
C6	14	I-2	We willen de data in de infrastructuur on premise hebben terwijl we voor de toepassingen software in de cloud willen gebruiken. Dat schuurt want dan moet je de brug tussen cloud en on-premise overbruggen. Manueel is dit niet gewenste maar in de automatisering is dat nog een uitdaging	Cloud Deployment Model
C6	15	I-2	Cloud ondersteuning zou zodanig moeten zijn dat je als klant tevreden bent en je technische randvoorwaarden stelt	Doelstelling Organisatie - Doelstelling CC
C6	16	I-2	Op dit moment worden vooral geleid door de leverancierskeuze (Microsoft Azure) We maken echter ook gebruik van private clouds. In de aanbestedingen wordt MS Azure niet als harde eis geformuleerd waardoor Euro's zodanig doorslaggevend zijn dat soms gekozen wordt voor een private cloud oplossing	Doelstelling Organisatie - Doelstelling CC
C7	17	I-1	Vanuit oogpunt van effectiviteit en efficiency kan men besluiten onderdelen van de Big Data uit te besteden.	Doelstelling Organisatie - Toe te passen CSM/CDM

Figuur 36: Resultaten wijze waarop CC voor BDA ingrijpt op EA – deel 1

Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
C8	18	I-1	De vraag is echter hoe ver men hier in wil gaan en hoe dat vervolgens aansluit op de invulling van de Business laag, Information Systems laag en Infrastructure laag van een EA van een organisatie. Hoe meer de provider in zijn Cloud Service Model opschuift richting de business laag – hoe minder controle wij nog kunnen uitoefenen.	CSM - Effect op EA lagen - Datakwaliteit
C9	19	I-1	Als je meer uitbesteed kun je als organisatie minder controle uitoefenen op de datakwaliteit. Waar leg je de grens en hoe hou je grip op de kwaliteit van je data?	CSM - Effect op EA lagen - Datakwaliteit
C9	20	I-1	Door te kiezen voor een hybride Cloud Deployment Model beperk je de afhankelijkheid en kun kwaliteitsrisico's beter mitigeren Als je kiest voor private hou je alles in eigen hand maar dat gaat misschien ten koste van andere doelstellingen. Zo kun je de Cloud bijvoorbeeld inzetten voor gedistribueerde verwerking wat de snelheid van het BDA proces ten goede komt.	CDM - Effect op EA lagen - Datakwaliteit
C9	21	I-2	Eisen ten aanzien van datakwaliteit moeten bij de introductie van CC meegenomen moet worden in de SLA's met de cloud provider	Datakwaliteit - Inzet CC

Figuur 37: Resultaten wijze waarop een CC voor BDA ingrijpt op een EA

### Samenvatting analyse resultaten

Onderstaande tabel geeft een samenvatting van de analyse resultaten waarbij, over de tekstfragmenten heen, het aantal keer is geteld dat een term of relatie tussen termen is voorgekomen. Het aantal weerspiegelt daarbij een mate van herkenbaarheid van een term of een relatie.

Term of relatie tussen termen	Aantal
Big Data Pipeline	2
Doelstelling Organisatie - Doelstelling CC	8
Doelstelling Organisatie - Toe te passen Cloud Service en Cloud Deployment Model	2
Doelstelling CC	1
Cloud Service Model - Effect op EA lagen - Datakwaliteit	2
Cloud Deployment Model - Effect op EA lagen - Datakwaliteit	1
Datakwaliteit	2
DataKwaliteit - Inzet CC	1
Cloud Deployment Model	1
Data Integratie	1

Figuur 38: Samenvatting analyse resultaten wijze waarop CC voor BDA ingrijpt op een EA



## Bijlage V: Resultaten inzet EAM voor impactbepaling bij migratie CC

Deze bijlage geeft voor de wijze waarop EAM kan worden ingezet voor impactbepaling bij migratie naar CC inzicht in:

- Onderkende termen vanuit de theoretische verkenning
- Geformuleerde vragen die als leidraad gold voor de gevoerde interviews
- Resultaten uit de gevoerde interviews en daaraan gekoppelde analyse
- Samenvatting van de analyse resultaten

### Onderkende termen

O.b.v. de verkregen inzichten vanuit een theoretisch oogpunt zoals weergegeven in sectie 4.5.1 zijn de volgende termen te onderkennen.

Termen
AS-IS situatie
Business Laag
Information Systems Laag
Infrastructuur Laag
Cloud Service Model (CSM)
Cloud Deployment Model (CDM)
TO-BE situatie
Roadmap
Projecten

Figuur 39: Termen EAM als instrument voor impactbepaling bij migratie naar CC

### Geformuleerde vragen

Op basis van het verkregen theoretisch inzicht en daaraan gekoppelde termen zijn onderstaande vragen geformuleerd.

Nr. Vraag	Vraag
D1	Op welke wijze wordt op dit moment EAM ingezet voor impactbepaling en doorvoeren van veranderingen en hoe wordt dit ervaren
D2	Op welke manier is de huidige inrichting voor BDA op dit moment vormgegeven, vertaald naar componenten/lagen binnen een EA
D3	Op welke wijze is DQM hierin tot uitdrukking gebracht
D4	Op welke manier is een beeld geschetst van de toekomstige situatie voor BDA ondersteund middels CC en hoe is dit vertaald naar componenten/lagen van een EA. Voor zover hierover nog niet is nagedacht: Hoe wordt de totstandkoming van deze toekomstige situatie gezien
D5	Op welke wijze is hierin de toekomstige inrichting voor DQM meegenomen of ziet men hoe DQM hierin meegenomen kan worden
D6	Hoe wordt het toepassen van EAM gezien bij migratie naar de Cloud in zijn algemeenheid
D7	Hoe ziet men het bepalen van de impact door vergelijking van de AS-IS met de TO-BE situatie
D8	Waarom moet de beschrijving van de AS-IS en TO-BE situatie voldoen wil deze voldoende basis vormen voor impactbepaling
D9	Op welke wijze vertaald zich dit naar een portfolio van uit te voeren projecten
D10	Hoe wordt EAM ingezet voor monitoring van de uit te voeren projecten en borgen van de kwaliteit van het projectresultaat in lijn met de geschetste TO-BE situatie

Figuur 40: Vragen EAM als instrument voor impactbepaling bij migratie naar CC

### Resultaten uitgevoerde interviews en daaraan gekoppelde analyses

In onderstaande tabel zijn de antwoorden op de gestelde vragen weergegeven waarbij o.b.v. de tekst is gekeken aan welk term of relatie van termen dit is te relateren. Een antwoord kan hierbij opgeknippt zijn in meerdere tekstfragmenten om het vervolgens te kunnen relateren aan de termen en onderlinge relaties. Een groot deel van de termen en onderlinge relaties hebben hun oorsprong in het verkregen theoretisch inzicht. Een deel van de termen en onderlinge relaties kunnen aanvullend geïntroduceerd zijn vanuit de gegeven antwoorden.

Toelichting kolommen uit de tabel:

- Nr. Vraag: Link naar de tabel met geformuleerde vragen (het nummer correspondeert met het nummer in de tabel met vragen)
- Nr. Antwoord: Identificatie van een tekstfragment van het antwoord op een vraag
- Interview: Identificatie van het interview
- Tekst: Tekstfragment van een antwoord wat te koppelen is aan een term of relatie tussen termen
- Term of relatie tussen termen: Termen of relatie tussen termen die herkend kunnen worden vanuit het tekstfragment (bij relatie tussen termen zijn de termen gescheiden door een “-”)

Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
D1	1	I-3	Voor projecten werken we de IST en SOLL situatie uit (bijvoorbeeld bij een personeelsinformatiesysteem)	Adoptie EAM; Adoptie EA
D1	2	I-3	Architectuur-wise/TOGAF-wise werken is nooit van de grond gekomen vanwege tijdgebrek bij de beschikbare architecten	Adoptie EAM
D1	3	I-3	Je wilt de architectuur eens in de zoveel tijd toetsen en waar nodig ingrijpen zoals dit in de TOGAF cyclus is onderkend	Adoptie EAM
D1	4	I-4	Niet betrokken bij de hantering van EAM als methodiek. Vraag is ook of het voor de informatiespecialist relevant is	Adoptie EAM
D1	5	I-4	Ziet wel het voordeel van het in kaart brengen van de IST situatie en de SOLL situatie vertaald naar de lagen van een EA. Zeker als communicatiemiddel om de neuzen te richten	Adoptie EAM - Impact
D1	6	I-7	Geen idee of dit op dit moment EAM gebruikt wordt. In het geval dat het gebruikt wordt dient dit beter gecommuniceerd te worden	Adoptie EAM
D1	7	I-7	Opletten dat je een methodiek niet strikt volgens het boekje uitvoert maar op maat maakt kijkend naar de eigenschappen/eenheden van de organisatie	Adoptie EAM
D1	8	I-7	We moeten op zoek naar handvatten/houvast bij het doorvoeren van	Adoptie EAM
D1	9	I-7	Er moet nog energie gestoken worden in het overtuigen van de organisatie dat een Enterprise Architecture houvast biedt	Adoptie EA
D1	10	I-8	We hebben een projectbrief (in het verleden een projectvoorstel) dat goedgekeurd moet worden	Projecten - Projectbrief
D1	11	I-8	Na goedkeuring wordt een project opgestart	Projecten - Projectbrief - Goedkeuring
D1	12	I-8	Dit begint met (parallel) een projectplan en een PSA.	Projecten - Projectbrief - PSA; ProjectPlan
D1	13	I-8	De PSA biedt houvast bij het voorbereiden van aanbestedingen (we zijn zelf geen softwarebouwers). Dit kan zich vertalen naar een standaard systeem of een maatwerk systeem. Dit laatste liever niet maar je komt ontkomt er soms niet aan.	Projecten - PSA
D1	14	I-8	In de PSA wordt per vlak (uitgaande van 9 vlakken en de 2 aanvullende vlakken) aangegeven per principe wat de impact is. Hierin zitten ook principes in die gerelateerd zijn aan de datakwaliteit. Het is een soort checklist	Projecten - PSA - DQM
D1	15	I-8	Bij het voorbereiden van aanbestedingen kan de PSA gebruikt worden voor het formuleren van een programma van eisen (PvE). Bij meervoudige aanbestedingen kunnen je vervolgens de diverse aanbiedingen op het PvE/PSA toetsen. In geval van afwijkingen kun je PSA dan ook als leidraad gebruiken.	Projecten - PSA - PvE
D1	16	I-8	Een principe in het kader van de PSA zorgt bovendien voor bewustzijn ten aanzien van allerlei aspecten die spelen bij een implementatie of migratie. Je gaat erover nadenken hoe deze geborgd worden en je gaat kijken naar consequenties	Projecten - PSA
D2	17	I-3	Toegevoegde waarde van een EA voor invoering van Big Data Analytics wordt sterk betwijfeld daar het merendeel zich op de Information Systems laag afspeelt. Big Data Analytics is een middel op een organisatie doel te bereiken	Adoptie EA - BDA
D2/D3	18	I-7	Gegevensmanagement is niet meegenomen in een IST EA. Wel zijn er architectuurprincipes die gelden voor gegevensmanagement zoals data los van applicaties. Vraag is of die vanuit architectuur zijn ingegeven of vanuit logisch nadenken.	Adoptie EA - IST
D2/D3	19	I-7	Gegevensmanagement is organische gegroeide gegeven de mensen en de ingezette applicaties	Adoptie EA - IST
D2/D3	20	I-7	Behoeft die we vanuit Gegevensmanagement hadden, zijn vertaald naar projecten die vervolgens tot uitvoering zijn gebracht zonder een grondslag in een EA	Adoptie EA - IST
D2/D3	21	I-7	EA van IST situatie kan weldegelijk toegevoegde waarde hebben: oAls je ergens naar toe wilt moet je weten waar je vandaan komt oEr is geen sprake van een groene weid oTransformatie vindt plaats terwijl de winkel open moet blijven en dat maakt het moeilijk. oDat wil overigens niet zeggen dat alles fout is in een IST situatie. En je moet pas de IST situatie verlaten als de SOLL situatie zich bewezen heeft.	Adoptie EA - IST

Figuur 41: Resultaten EAM als instrument voor impactbepaling bij migratie naar CC - deel 1



Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
D3	22	I-3	Ook DQM ligt voor het merendeel op de Information Systems laag en niet op de gehele EA. Ook hier wordt de toegevoegde waarde van een EA sterk betwijfeld.	Adoptie EA - DQM
D3	23	I-7	Data Quality Management zou hierin meegenomen moeten worden. Data staat centraal. Organisatie moet aandacht besteden aan hoge data kwaliteit vanwege hun taken en de vereiste transparantie hoe we omgaan met data. Dit zou een organisatie-breed vraagstuk moeten zijn met bijbehorende bewustwording. Tot nu toe wordt het nog te zeer vanuit Gegevensmanagement gedreven. Wil je als organisatie datgedreven willen werken moet je voorzien in uitgangspunten/principes die gelden ten aanzien van datakwaliteit en daar uitvoering aan geven.	Adoptie EA - DQM
D4	24	I-3	We hebben een document opgesteld m.b.t. sourcing strategie waarbij doelstellingen van de organisatie zijn doorvertaald naar cloudvarianten. Inclusief duiding waar willen als organisatie naar toe willen. Interessant document. Goedkeuring van dit document is echter uitgebleven	Doelstelling Organisatie - Doelstelling CC
D4	25	I-3	In de tussentijd op applicatieniveau werken wel toe naar SaaS oplossingen maar niet op infrastructureel niveau.	Toepassing CC - SOLL - EA lagen
D4	26	I-3	Als ik echter SaaS oplossingen introduceer breng ik wel bepaalde infrastructurele onderdelen zoals opslag (voor die specifieke toepassing) wel naar de Cloud. Dus gaandeweg wordt de infrastructuur uitgedund.	Toepassing CC - SOLL - EA Lagen
D4	27	I-3	Cloud applicaties worden verweven in onze architectuur en we stellen concrete eisen aan deze applicaties. Bijv. Exit strategie; datacenter in Europa, veiligheid, service-oriëntatie etc.	Eisen aan CC
D4	28	I-3	Reden om toch naar de Cloud te gaan wordt veelal ingegeven door de markt. Bijna alle hedendaagse toepassingen zijn Cloud toepassingen. On premise komt bijna niet meer voor	Toepassing CC
D4	29	I-8	Cloud Service en Deployment Model en daarin onderkende niveaus zijn herkenbaar	Toepassing CC
D4	30	I-8	Leveranciers bieden toepassingen vaak bijvoorbeeld alleen als een SaaS oplossing aan.	Toepassing CC - Leveranciers
D4	31	I-8	Hetzelfde geldt voor partijen als Microsoft die meer richting CC gaan.	Toepassing CC - Leveranciers
D4	32	I-8	Je wordt vanuit markt gedwongen naar de Cloud te gaan. Er zitten wel wat zorgpunten aan vast. Denk aan informatieveiligheid en de wijze waarop je het beheer moet inrichten.	Toepassing CC - Leveranciers - Infobeveiliging & Beheer
D4	33	I-8	Ook koppelvlakken is dan een aandachtspunt. Daarvoor hebben we een voorziening als een Enterprise Service Bus geïntroduceerd om koppelingen veilig in te richten.	Toepassing CC - ESB
D4	34	I-8	De koninklijke weg is om vanuit de organisatiedoelstellingen te kijken naar bijbehorend cloudstrategie.	Doelstelling Organisatie - Doelstelling CC
D4	35	I-8	We hebben, vanuit de afdeling op eigen initiatief, in het verleden een document voor een cloud en sourcing strategie opgesteld met een bijbehorende routekaart. Gegeven ook ontwikkelingen in de capaciteit en	Doelstelling Organisatie - Doelstelling CC
D4	36	I-8	Ondanks dat dit document geen vervolg heeft gekregen is het wel een interessant document. Het is echter wel niet duidelijk of de teammanagers	Doelstelling Organisatie - Doelstelling CC
D4	37	I-8	Op dit moment is niet echt een lijn in de migratie naar CC te zien zoals onderkend in het eerder genoemd document. Op dit moment is het meer ad hoc georganiseerd.	Toepassing CC
D4	38	I-8	De gehanteerde methodiek om te komen tot een cloud/sourcing strategie zoals verwoord in het eerder genoemd document wordt wel onderschreven (los dat het geen vervolg heeft gekregen). De waarom, wie, wat, hoe, wanneer vragen worden gesteld. Het heeft haakjes naar doelstellingen waarbij ook rekening wordt gehouden met de verscheidenheid aan doelstellingen per team. Sessies waren intensief waarin diverse disciplines zijn betrokken.	Doelstelling Organisatie - Doelstelling CC
D4/D5	39	I-4	Voor keuze van het toe te passen Cloud Service Model (IaaS, PaaS, SaaS etc.) spelen de volgende aspecten: - Kosten (waarbij je in geval van CC alleen betaald voor hetgeen je afneemt) - Vertrouwelijkheid en informatieveiligheidsaspecten - Ruimte en noodzakelijke bemensing (middels CC kun je de organisatie ontzorgen zowel voor de benodigde expertises als de noodzakelijke ruimte die je nodig hebt voor je infrastructuur)	Doelstelling Organisatie - Doelstelling CC
D4/D5	40	I-4	Ten aanzien van het Cloud Deployment Model (private, public, hybrid) wordt aangegeven dat aspecten die hier en rol spelen zijn: - Mate van vertrouwelijkheid die geldt (kiezen voor private waar een hoge mate van vertrouwelijkheid geldt) - Wettelijke bepalingen die gelden en bijbehorende veiligheidsaspecten - Eisen vanuit de ketenpartners	Doelstelling Organisatie - Doelstelling CC
D4/D5	41	I-4	Ten aanzien van de vertaling naar een SOLL situatie wordt aangegeven dat dit best complex is gezien de mogelijke varianten in het Cloud Service Model en Cloud Deployment Model. Je zou kunnen beginnen met het uitwerken van een drietal varianten met een wezenlijk verschil tussen de varianten	Toepassing CC - SOLL - Lagen EA
D4/D5	42	I-7	Als gegevensmanagement moeten wij ons werk goed doen Hiervoor stellen we randvoorwaarden aan de data, het werken met data en bijbehorende informatieveiligheidsvraagstukken. Wij willen hierin geadviseerd worden. Wij willen niet verantwoordelijk zijn voor de ICT infrastructuur	Doelstelling Organisatie - Doelstelling CC

Figuur 42: Resultaten EAM als instrument voor impactbepaling bij migratie naar CC - deel 2

Nr. Vraag	Nr. Antwoord	Interview	Tekst	Term of Relatie tussen termen
D4/D5	43	I-7	Wijze waarop dit vervolgens ondersteund wordt (on premise of CC of combinatie) is niet ons vraagstuk. Wij moeten ontzorgd worden in het uitvoeren van onze werkzaamheden, gegeven de randvoorwaarden die gelden. Als daar CC een oplossing in biedt: prima zolang we er geen nadelen van ondervinden. Bijvoorbeeld infrastructurele hobbels waar we nu mee geconfronteerd worden. Daar willen we geen last van hebben.	Doelstelling Organisatie - Doelstelling CC
D4/D5	44	I-7	Architect dient, uitgaande van de eisen/randvoorwaarden die wij stellen, de vertaalslag maken naar een duurzame oplossing (raamwerk) met inzet van CC (incl. bijbehorende consequenties) waarbij er ruimte is om ook in de toekomst ons werk te kunnen doen, rekening houdend met de dan geldende technologische standaarden. Architectuur moet niet alles in kleuren.	Toepassing CC - SOLL - Lagen EA
D4/D5	45	I-7	Dat betekent dat er sprake moet zijn van architectuurprincipes die niet gekoppeld zijn aan bepaalde technologische ontwikkelingen en dat je kunt migreren naar een volgend niveau, gegeven de stand van de technologie op dat moment.	Adoptie EA
D4/D5	46	I-7	Belangrijk is dat de Architect en elkaars taal begrijpen en dat er ook gekeken wordt naar mogelijke alternatieven, gegeven de haalbaarheid. Uiteindelijk moet er een advies komen gegeven de mogelijke alternatieven.	Adoptie EA
D4/D5	47	I-7	Bij het komen tot een SOLL situatie moeten we goed voor ogen hebben waar we wel en niets in willen betekenen, gegeven de doelstellingen waar de organisatie voor staat en rekening houdend met aspecten op het gebied integriteit en ethiek. We hoeven immers niet alles te adopteren wat vanuit de markt gepromoot wordt. Dus ook aangeven waarin je niet meegaat.	Adoptie EA - SOLL
D4/D5	48	I-7	De architectuur dient rekenschap te houden met de afhankelijkheden tussen de onderdelen van een EA en terugvalscenario's te hebben voor de kritische onderdelen.	Adoptie EA
D4/D5	49	I-7	De architectuur moet in een groter geheel geplaatst worden zodat ook de voordelen voor de gehele organisatie duidelijk worden.	Adoptie EA
D5	50	I-7	DQM aspecten moeten hierin meegenomen worden. Waarbij ook de motivatie (het waarom) van DQM wordt meegenomen. Je moet de organisatie overtuigen van de voordelen hierin en aangeven welke risico's we lopen als we er geen invulling aan geven. We moeten aangeven wat wel doen maar zeker ook aangeven wat we niet gaan doen met bijbehorende consequenties. Zoals eerder aangegeven is het van belang dat dit een organisatie-breed vraagstuk is, niet beperkt tot Gegevensmanagement. Het gedachtegoed moet organisatie-breed gedragen worden	Adoptie EA - DQM
D6	51	I-7	Aandacht besteden aan de zachte kant betekent medewerkers overtuigen van de voordelen van de migratie en hiervoor vertrouwen winnen (er speelt immers een onzekere factor voor de mensen). De veranderkundige aspecten moeten niet onderschat worden en vergen een flinke inspanning.	Adoptie EAM - Impact - Veranderkundige aspecten
D6	52	I-7	In geval EAM geen aandacht besteedt aan de zachte (veranderkundige) kant zijn er onvoldoende waarborgen om de transitie/migratie tot een succes te	Adoptie EAM - Impact - Veranderkundige aspecten
D9	53	I-7	Bij het vertalen van de IST naar de SOLL in een roadmap moet niet alleen aandacht wordt besteed aan de harde (instrumentele) kant maar moet ook aandacht wordt besteed aan de zachte kant waarbij mensen dingen moeten afleren en nieuwe dingen moeten aanleren. Het zijn immers dezelfde	Adoptie EAM - Roadmap - Veranderkundige aspecten
D9	54	I-7	We moeten voorzien in geleidelijke migraties waarbij kleine resultaten boeken die uiteindelijk bijdragen in het einddoel, gegeven de huidige organisatie en het verandervermogen dat de organisatie heeft. Je hebt immers met een grote organisatie te maken waarbij de wendbaarheid beperkingen heeft. Daarbij moet in het oog worden gehouden dat het	Adoptie EAM - Roadmap - Veranderkundige aspecten
D10	55	I-7	Gegeven de complexiteit van een transitie/migratie is monitoring noodzakelijk.	Adoptie EAM - Monitoring
D10	56	I-7	Wordt in de tussenresultaten nog voldoen aan de eisen en randvoorwaarden die ik van tevoren heb meegegeven.	Adoptie EAM - Monitoring
D10	57	I-7	Deze monitoring moet preventief en anticiperend zijn en niet reactief.	Adoptie EAM - Monitoring
D10	58	I-7	Monitoring kan leiden tot bijstelling van projecten maar kan ook leiden tot bijstelling van de SOLL architectuur. Deze flexibiliteit moet wel aanwezig zijn.	Adoptie EAM - Monitoring

Figuur 43: Resultaten EAM als instrument voor impactbepaling bij migratie naar CC – deel 3

### *Samenvatting analyse resultaten*

Onderstaande tabel geeft een samenvatting van de analyse resultaten waarbij, over de tekstfragmenten heen, het aantal keer is geteld dat een term of relatie tussen termen is voorgekomen. Het aantal weerspiegelt daarbij een mate van herkenbaarheid van een term of relatie.

Term of relatie tussen termen	Aantal
Adoptie EAM Algemeen	6
Adoptie EAM - Veranderkundig	2
Adoptie EAM - Impact - Veranderkundige aspecten	2
Adoptie EAM - Roadmap - Veranderkundige aspecten	2
Adoptie EAM - Monitoring	4
Adoptie EA Algemeen	6
Adoptie EA - DQM	2
Adoptie EA - BDA	1
Adoptie EA - Weergave ASIS	4
Adoptie EA - Weergave TOBE	1
Projecten - Project Start Architectuur	2
Projecten - Project Start Architectuur - DQM	1
Projecten - Project Start Architectuur - Programma van Eisen	1
Projecten - ProjectBrief	1
Projecten - ProjectBrief - Goedkeuring	1
Projecten - ProjectBrief - Project Start Architectuur/Projectplan	1
Doelstelling Organisatie - Doelstelling CC	9
Eisen CC	1
Toepassing CC	3
Toepassing CC - TOBE - EA Lagen	4
Toepassing CC - Leveranciers	2
Toepassing CC - Leveranciers - Infoveiliging & Beheer	1

*Figuur 44: Samenvatting analyse resultaten EAM als instrument voor impactbepaling bij migratie naar CC*